

STIC-Biotech/ChemLib

174948

From: Ramirez, Delia
Sent: Thursday, December 22, 2005 4:38 PM
To: STIC-Biotech/ChemLib
Subject: 10/602219

Hi,

I would like to request the following searches:

1. SEQ ID NO: 11 and 12 in the nucleic acid databases (commercial and interference)
2. SEQ ID NO:12 in the protein databases (commercial and interference)

Thank you very much,

Delia M. Ramirez, Ph.D.
Patent Examiner
Recombinant Enzymes-Art Unit 1652
USPTO
400 Dulany Street, Remsen Bldg., 2D74, Mail room 2C70
Alexandria, VA 22314
(571) 272-0938
delia.ramirez@uspto.gov

RECEIVED
DEC 22 2005
1517

Searcher: DiNatale
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 10:52:16 ; Search time 233 Seconds
(without alignments)
3211.817 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MOLRNPELHLCALALRFLA.....RLRSHINPTGVLLQLENTM 421

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool.p/US10602219/runat.23122005.151142.17165/app.query.fasta_1.583
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MAIRX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602219 @CNC 1.1.290 @runat.23122005.151142.17165 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2.6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2.6/ptodata/1/ina/H COMB.seq.*
6: /cgn2.6/ptodata/1/ina/PCITUS COMB.seq.*
7: /cgn2.6/ptodata/1/ina/pp COMB.seq.*
8: /cgn2.6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	1266	3	US-09-993-059-11
2	2314	100.0	1266	3	US-10-103-327-11
3	2314	100.0	1278	3	US-09-993-059-7
4	2314	100.0	1278	3	US-10-103-327-7
5	2314	100.0	1284	3	US-09-993-059-13
6	2314	100.0	1284	3	US-10-103-327-13
7	2314	100.0	1290	3	US-09-993-059-3
8	2314	100.0	1290	3	US-10-103-327-3
9	2314	100.0	1296	3	US-09-993-059-9

10	2314	100.0	1296	3	US-10-103-327-9	Sequence 9, Appli
11	2314	100.0	1308	3	US-09-993-059-5	Sequence 5, Appli
12	2314	100.0	1308	3	US-10-103-327-5	Sequence 5, Appli
13	2314	100.0	1343	3	US-08-928-881-18	Sequence 18, Appl
14	2314	100.0	1343	3	US-09-543-921-18	Sequence 18, Appl
15	2314	100.0	1343	3	US-09-266-014-3	Sequence 3, Appli
16	2314	100.0	1343	3	US-09-491-759-18	Sequence 18, Appl
17	2314	100.0	1393	2	US-07-602-824A-1	Sequence 1, Appli
18	2314	100.0	1393	2	US-07-983-451-1	Sequence 1, Appli
19	2314	100.0	1393	2	US-08-261-577-6	Sequence 6, Appli
20	2293	99.1	1254	3	US-09-993-059-15	Sequence 15, Appl
21	2293	99.1	1254	3	US-10-103-327-15	Sequence 15, Appl
22	2293	99.1	1272	3	US-09-993-059-17	Sequence 17, Appl
23	2293	99.1	1272	3	US-10-103-327-17	Sequence 17, Appl
24	2293	99.1	8234	3	US-09-626-127-14	Sequence 14, Appl
25	2293	99.1	8234	3	US-09-993-059-34	Sequence 34, Appl
26	2293	99.1	8234	3	US-10-103-327-34	Sequence 34, Appl
27	2234	96.5	1304	9	5179023-3	Patent No. 5179023
28	2231	96.4	1233	3	US-09-993-059-21	Sequence 21, Appl
29	2231	96.4	1233	3	US-10-103-327-21	Sequence 21, Appl
30	2227	96.2	1215	3	US-09-993-059-19	Sequence 19, Appl
31	2227	96.2	1215	3	US-10-103-327-19	Sequence 19, Appl
32	2178	94.1	1214	3	US-09-023-655-1001	Sequence 1001, Ap
33	2152	93.0	1170	3	US-09-176-666-49	Sequence 49, Appl
34	2152	93.0	1173	3	US-09-176-666-48	Sequence 48, Appl
35	2152	93.0	1176	3	US-09-176-666-47	Sequence 47, Appl
36	2152	93.0	1179	3	US-09-176-666-46	Sequence 46, Appl
37	2152	93.0	1182	3	US-09-176-666-45	Sequence 45, Appl
38	2152	93.0	1188	3	US-09-176-666-44	Sequence 44, Appl
39	2152	93.0	1194	3	US-09-176-666-43	Sequence 43, Appl
40	2152	93.0	1197	3	US-08-928-881-25	Sequence 25, Appl
41	2152	93.0	1197	3	US-09-543-921-25	Sequence 25, Appl
42	2152	93.0	1197	3	US-09-266-014-5	Sequence 5, Appli
43	2152	93.0	1197	3	US-09-491-759-25	Sequence 25, Appl
44	2147	92.8	1167	3	US-09-176-666-50	Sequence 50, Appl
45	2142	92.6	1164	3	US-09-176-666-51	Sequence 51, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-11
; Sequence 11, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUNAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-993-059-11

Alignment Scores:	8.08e-252	Length:	1266
Pred. No.:	2314.00	Matches:	421
Score:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-10-602-219-12 (1-421) x US-09-993-059-11 (1-1266)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTTCGGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCCTGGGCTAGAGCACTGGACAATGATGGCAAGGAGCCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGACCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTCGCATCAGTGAAGGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGGATGCAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAAGAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATTATGTTCAACAGCAAGAGACTGAAGCTAGGAGTTATGCAAGTTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCTGGAGTTTGGATACCTACGACATTGATGCCCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGlnLeu 180
DB 481 GACTGGGAGTAGACTGCTGAAAATTTGATGGTGTACTGTGACACTTTGGAAAATTGG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATGAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAACATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAAATTTTGCTGCAATTGATGATTCCTCGGAAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGACTGGACATCTTTTACCAGGAGAAATTTGTGATGTGCTGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGAAATGACCCAGATATGTTAGTGTGTAATGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATGAAGGAGCTTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGCAGCAAGGGTACCACCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCGCTGGCTGTAGTATGATAAACCAGCGGAGAGATTGGT 1080
```

```
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGlnTrpThr 400
DB 1141 GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
QY 421 Met 421
DB 1261 ATG 1263
RESULT 2
US-10-103-327-11
; Sequence 11, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 08010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-103-327-11
Alignment Scores:
Pred. No.: 8,08e-252 Length: 1266
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-10-602-219-12 (1-421) x US-10-103-327-11 (1-1266)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTTCGGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCCTGGGCTAGAGCACTGGACAATGATGGCAAGGAGCCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGACCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTCGCATCAGTGAAGGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGGATGCAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
```



```
Db 301 GATTTCAGAGGAGAGCTTCAGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAntyValHisSerLysGlyLeuLysLeuGlyLysLeuValHisSerLys 140
Db 361 GCTAATTAATGTTACAGCAGAGGAGCTGAGCTAGGAGATTAAGCATGTTGGAAATANA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCTGGGAGTTTGGATACATGACATTCAGCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTCTCTAAATTTGATGGTTTGTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACAATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGCCCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTAACAGAAATCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTCATGATTCCTGGAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGAGCTGACATCTTTTAAACAGGAGAGAAATTTGTGATGTTGCTGGACAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaLalaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTAAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCCAAGCTCTCTTCAGGATAAGGAGTAAATTCGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGACGACGAGGTACAGCTTAGACAGGAGAGACAACTTTGAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGAGCTATGATAAACCAGCAGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaLaserLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCCAGTCCAGTTCCTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCTGTAAGAGAGAGCTAGGGTTTCTATGATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAAACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
```

RESULT 3

```
US-09-993-059-7
; Sequence 7, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
```

```
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-09-993-059-7

Alignment Scores:
Pred. No.: 8,19e-252 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
```

US-10-602-219-12 (1-421) x US-09-993-059-7 (1-1278)

```
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuAtpPheLeuAla 20
Db 1 ATGACGCTGAGAACCCAGAACATCATCTGGGCTGGCGCTTGGCGCTTCGCTTCCTCGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTCTGGGACATCCCTGGGCTAGACATCGACATGATGATGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGTCAGAGAGGCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGACAGATTAAGTACTCTGCAATGATGACTGTTGGATGGCTCCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGACAGACTTCAGGACAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAntyValHisSerLysGlyLeuLysLeuGlyLysLeuValHisSerLys 140
Db 361 GCTAATTAATGTTACAGCAGAGGAGCTGAGCTAGGAGATTAAGCATGTTGGAAATANA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCTGGGAGTTTGGATACATGACATTCAGCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTCTCTAAATTTGATGGTTTGTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACAATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGCCCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTAACAGAAATCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
```

Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTTGATGATTCCTGGAAAAAGTATAAAG 720
QY 241 SerIleLeuAspTrrThrSerPheAsnGlnGluArgIleValAspValalaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACAGGG 780
QY 261 GlyTrrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrrPasnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCTCAGCTGGAAATCAGCA 840
QY 281 ValThrGlnMetAlaLeuTrrPalaIleMetAlaLaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrrGlnLeuArgGlnGlyAspAsnPheGluValTrr 340
Db 961 CAGGACCCCTTGGCCAGCAAGGTTACCACTTACAGGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrrPalaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAAGGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGGCAGGAGATTGCT 1080
QY 361 GlyProArgSerTrrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrrGluTrrThr 400
Db 1141 GCTGTCTTATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAGTCACATAATCCACAGGCACCTGTTTGTCTTCAGCTAGAAAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263

RESULT 4

US-10-103-327-7
; Sequence 7, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-103-327-7

Alignment Scores:
Pred. No.: 8,19e-252 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-10-602-219-12 (1-421) x US-10-103-327-7 (1-1278)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGGCGCTTGGCTTCGCTTCCTCGCC 60
QY 21 LeuValSerTrrPaspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCCTGGCAATGGATTTGGCAAGGACGCT 120
QY 41 ThrMetGlyTrrLeuHisTrrPaspIleProGlyAlaArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCTGGAGGCTTCATGTCAACCTTTGATGCGAGGAAGGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrrLysAspAlaGlyTrrGluTrrLeuCysIleAspAspCysTrrPaspIleProGlnArg 100
Db 241 TGGAAAGGATGCAGGTTATGAGTACCTCTGCATTTGATGACTGTGTGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGCAGACCTTCAGCAGACCTTCAGCGCTTTCTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTrrValHisSerLysGlyLeuLysLeuGlyIleTrrValAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrrTrrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTCGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTTGATGCGCAGACCTTTGT 480
QY 161 AspTrrGlyValAspLeuLysPheAspGlyCysTrrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTTGTGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTrrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrr 200
Db 541 GCAGATGTTTATAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrrProLeuTrrMetTrrProPheGlnLysProAsnTrrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTrrCysAsnHisTrrArgAsnPheAlaAspIleAspAspSerTrrLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCACTTGTATGATTCCTGGAAAGTATAAG 720
QY 241 SerIleLeuAspTrrThrSerPheAsnGlnGluArgIleValAspValalaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACAGGG 780
QY 261 GlyTrrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrrPasnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrrPalaIleMetAlaLaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrrGlnLeuArgGlnGlyAspAsnPheGluValTrr 340
Db 961 CAGGACCCCTTGGCCAGCAAGGTTACCACTTACAGGGAGACAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeuGly 360
 DB 1021 GAACGACCTCTCTAGGCTTAGCCTGGCTGTAGCTATGATAAACCAGCAGGAGATTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 DB 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAAGAAAGCTAGGTTTCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 DB 1201 TCAAGTTTAAAGAGTCACATAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260
 QY 421 Met 421
 DB 1261 ATG 1263
 RESULT 5
 US-09-993-059-13
 ; Sequence 13, Application US/09993059
 ; Patent No. 6887696
 ; GENERAL INFORMATION:
 ; APPLICANT: GARGER, Stephen A.
 ; APPLICANT: TURPEN, Thomas H.
 ; APPLICANT: KOMAGAI, Monto H.
 ; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
 ; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
 ; FILE REFERENCE: 008010087CPUS06
 ; CURRENT APPLICATION NUMBER: US/09/993,059
 ; CURRENT FILING DATE: 2001-11-13
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 1284
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1284)
 US-09-993-059-13
 Alignment Scores:
 Pred. No.: 8,25e-252 Length: 1284
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-10-602-219-12 (1-421) x US-09-993-059-13 (1-1284)
 QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTTCCTGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTCGTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGATGGCAAGGAGCGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTTCATGTGCAACTTTGACTGCCAGGAAGGCCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTCTGTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
 QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100

RESULT 6

US-10-103-327-13

; Sequence 13, Application US/10103327

DB 241 TGGAAGGATGAGGTTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTCAAGAGGACAGACTTCAGGACAGCCCTCAGCGCTTCTCTATGGGATTCGCCAGCTA 360
 QY 121 AlaIleTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 DB 361 GCTAATTAATGTTTACAGCAAGAGCTGAAGCTAGGAGTTTATGCAAGATGTTTGGAAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGGCGAGGCTTCCCTGGGAGTTTTGGATACACGACATTTGATGACAGTTTGGAAAATTG 480
 QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTG 540
 QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 DB 541 GCAGATGGTTATAAGCACATGCTTGGCCCTGTAATAGGACTGGCAGAGCAATTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 DB 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 DB 661 CAGTACTGCATCACTGGCGAAATTTTCTGCATGATGATTCCTGGAAAGATATAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGGACTGACATCTTTAAACCCAGGAGAGATTTGTGTGTTGCTGGACAGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
 DB 781 GGTGGAAATGACCCAGATATGTTAGTGTGTTGGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTCATGCTTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGACGTAAATTCGCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 DB 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGTATGATATAACCGCAGAGATTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 DB 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGAAAGCTAGGTTTCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 DB 1201 TCAAGTTTAAAGAGTCACATAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260
 QY 421 Met 421
 DB 1261 ATG 1263

```

; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103.327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-103-327-13

Alignment Scores:
Pred. No.:      8,25e-252      Length:      1284
Score:          2314.00      Matches:      421
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              3          Gaps:      0

US-10-602-219-12 (1-421) x US-10-103-327-13 (1-1284)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCGCTTCCTGGCC 60

Qy 21 LeuValSerTrpAspIleProGlyAlaAargAlaLeuAspIleGlyLeuAlaArgThrPro 40
Db 61 CTCGTGTTCTCGGACATCCCTGGGCTAGACACCTGAGCAATGGATTGGCAAGACGCCT 120

Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGGCTGGCTGCATCTGGAGGCGCTTCATGTGCACACTTGACTGCCAGGAAGACCA 180

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

Qy 81 TrpLysAspAlaGlyTyxGluTyrlieuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGAGATGCAAGGTATAGTACCTTCGCAATTGATGACTGTTGGATGGCTCCCCAAGA 300

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGGACGAGCTTCAGCGACACCTCAGGCGCTTCTCATGGGATTCGCCAGCTA 360

Qy 121 AlaAsnTyxValHiserLysGlyLeuLysLeuGlyIleTyxAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAAATAA 420

Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyxTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTTGGATACTACGACATTTGATGCCACAGCTTTGCT 480

Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyxTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTTCATGGTTGTACTGTGACAGTTTGGAANAATTG 540

Qy 181 AlaAspGlyTyxLysHiserLysLeuAlaLeuAsnArgThrGlyArgSerIleValTyx 200
Db 541 CGAGATTGGTTATAAGCATTGCTTCCTGGCCCTGAATAGGACTGGCAGAAGCATTTGTAC 600

```

Pred. No.: 8,31e-252 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-09-993-059-3 (1-1290)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGGCTTGGCTTCCTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCATCTGGGAGGCTTCATGTGCAACCTTGTGCTGCCAGGAGGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGCAGGTATGAGTACCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAGAAGGAGCAGCTTCAGGAGAGCTTCAGGCTTCCTCATCTGGGATTCGCCAGCTA 360

QY 121 AlaAsnTrpValHisSerIysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 361 GCTAATTTATGTCACAGCAAGAGCTGAAGCTAGGAGATTATGCAAGATGTTGGAATATAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTCCCTGGGAGTTTGGATACAGCACTGATGATGATGATGATGATGATGAT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerIysLeuAsnLeu 180
DB 481 CACTGGGAGTAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 541 GCAGATGGTTATAGCACATGCTTGGCCCTGGAATAGGACTGGCAGAGCACTTGTGTAC 600

QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
DB 601 TCTGTGAGTGCCCTTTATATGTCGCTTTCAGAGCCCAATATACAGAAATCCGA 660

QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGGAAATTTTGTGATGATGATGATGATGATGATGATGATGATGAT 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAAACCCAGGAGAAATTTGTTGATGTTGCTGGACAGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCTCTGGGCTATCATGGCTGCTCTCTTTTATTCATGTCTAATGACCTC 900

QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACATCAGGCCCTCAAGCCAAAGCTCTCTTCCAGGATAGGAGCTTAATTCCTCAAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTrpClnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGCACCCCTTTGGGCAAGAGGTTACAGCTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCCGAGGAGATTGGT 1080

QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTTATACCATCGCAGTTCCTCCCTGGGTAAGGAGTGGCCTGTAAATCCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGAGCTAGGCTTCTATGAATGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGTTAAGAGTCACTAATAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260

QY 421 Met 421
DB 1261 ATG 1263

RESULT 8
US-10-103-327-3
; Sequence 3, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-10-103-327-3

Alignment Scores:
Pred. No.: 8,31e-252 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-10-103-327-3 (1-1290)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGGCTTGGCTTCCTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCATCTGGGAGGCTTCATGTGCAACCTTGTGCTGCCAGGAGGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80

```
Db 181 GATTCTCGCATCAGTACAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGTCAGGTATAGTACCTCTCATTTGATGACTGTGTGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAGGAGACTTCAGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGCAATTTATGCCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAAATAGGACTGGCAGAGCAATTTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCGTGAGTGGCCCTCTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATACACTGGCGCAATTTTGTCTGCATTTGATGATCTCTGGAAAAGTATTAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnArgIleValAspValalaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACACGAGGAGAAATTTGTGATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetIleValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAATGACCCAGATATGTAGTATGGCACTTTGGCCCTCAGCTGGAAATCAGCA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTCATAGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGATACCAAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATTAACCCGCGAGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGAGTGGCCGTGAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValIleLeuGlnLeuGluAsnThr 420
Db 1201 TCAGGTTAAGAGTCACATTAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260
QY 421 Met 421
```

```
Db 1261 ATG 1263
RESULT 9
US-09-993-059-9
; Sequence 9, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUNAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1296)
US-09-993-059-9
Alignment Scores:
Pred. No.: 8,37e-252 Length: 1296
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-10-602-219-12 (1-421) x US-09-993-059-9 (1-1296)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGCTGCGGCTTGGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCAGTGGCAATGGATGGCAAGGACGCT 120
QY 41 ThrMetGlyTyrLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCATCTGGAGGCGCTTCATGTCAACCTTGACTGCCAGGAAGGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGTCAGGTATAGTACCTCTGCATTTGATGACTGTGTGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAGGAGACTTCAGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGCAATTTATGCCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
```



```
Db 541 CGAGATGGTTAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCCCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTTGATGATTCCTGGAAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAAATTTGTGTTGCTGGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATTTAGTGTATGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTATTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCCTCAATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGACAGCAGGATACAGCTTAGACAGGAGACACTTTGAAGTGGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGCGCAGGAGATTGCT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGAGTGGCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGlnTrpThr 400
Db 1141 GCCTGCTTATCACAGCTCTCTCCCTGTGAAAGAGAACTAGGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAGTACACATAATCCACAGGCACTGTTTGTCTCAGCTAGAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
```

RESULT 10

```
US-10-103-327-9
; Sequence 9, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Stephen A.
; APPLICANT: KOMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (1)....(1296)
US-10-103-327-9
Alignment Scores:
Pred. No.: 8,37e-252 Length: 1296
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
```

US-10-602-219-12 (1-421) x US-10-103-327-9 (1-1296)

```
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGGC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCATGTCAACCTTGACTGCCAGGAAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAGGATGAGGTTATGAGTACCTCTGCATTTGATGATGATGTTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAGGAGAGACTTCAGGACACCTTCAGCCCTTTCCTCATGGGATTGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGGATTATGCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTCGCAGGCTTCCCTGGGAGTTTGGATACACGACATTTGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAGCAATGCTCTTGGCCCTGATAGGACTGGCAGAGCAATTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTTGATGATTCCTGGAAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAAATTTGTGTTGCTGGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATTTAGTGTATGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTATTCATGTCTAATGACCTC 900
```

```
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAen 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGCGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263

RESULT 11
US-09-993-059-5
; Sequence 5, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1308)
US-09-993-059-5

Alignment Scores:
Pred. No.: 8,49e-252 Length: 1308
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-09-993-059-5 (1-1308)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTCTCGGACATCCCTGGGGCTAGAGCACTGGACATGGATTGGCAAGGAGCCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
```

```
Db 121 ACCATGGCGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTGCATCAGTCAGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TCGAAGGATCAGAGGTATAGTACCTCTGCAATTCATGACTGTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyLysLeuGlnLeu 120
Db 301 GATTTCAGAAAGCAGACCTTCAGCAGACCTTCAGCCCTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTACAGCAAGAGACTGAAGCTAGGGAATTCATGCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGAGATTTTGGATACTACGACATTCATGATGCCCAGACCTT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTTGTGTTGTTACTGTGACAGATTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAGCACATGCTCTTGGCCCTGANTAGACTGGCAGAAAGCAATTCGTATC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTCATGATTCCTCGAAAGATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTAAACAGGAGAGAAATTTGTGATGTTGCTGACACGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTTCGAATGACCCAGATATGTTAGTATTCGCAACTTTGGCCTTCAGCTGGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTCATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAen 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGCGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
```


QY 421 Met 421
 DB 1261 ATG 1263

RESULT 12

US-10-103-327-5
 ; Sequence 5, Application US/10103327
 ; Patent No. 6890748
 ; GENERAL INFORMATION:
 ; APPLICANT: GARGER, Stephen A.
 ; APPLICANT: TURPEN, Thomas H.
 ; APPLICANT: KUMAGAI, Monto H.
 ; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
 ; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
 ; FILE REFERENCE: 008010087CPUS06
 ; CURRENT APPLICATION NUMBER: US/10/103,327
 ; CURRENT FILING DATE: 2002-03-20
 ; PRIOR APPLICATION NUMBER: US/09/993,059
 ; PRIOR FILING DATE: 2001-11-13
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 1308
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1308)
 US-10-103-327-5

Alignment Scores:

Pred. No.: 8,49e-252 Length: 1308
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-10-103-327-5 (1-1308)

QY 1 MetGlnLeuArgAspProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGAGCTGAGGAAACCCAGAACTATCTGGGCTGCGGCTTGGCTTCCTGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGGATTGGCAAGAGGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGGCTGGCTGCACCTGGGAGGCTTCATGTGCAACCTTGTACCTGCCAGGAGCCA 180
 QY 61 AspSerCysIleSerGlnLeuLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TGAAGGATGCAAGTTATGAGTACTCTGCAATGATGCTGTGGATGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTTCAGAAGGCAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 QY 121 AlaAsnTrpValHisSerIleGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
 DB 361 GCTAATTTATGTTTCACAGCAAGAGCTGAAGCTAGGATTTATGCAGATGTTGGAAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGGCAGGCTTCCTCCCTGGGAGTTTGGATCTACGACATTTGATGCTCCAGACCTTTCCT 480

QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540
 QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 DB 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
 DB 601 TCCTGTAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspLeuAspSerTrpLysSerIleLys 240
 DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTTGATGATTCCTGGAAGATTAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGATTTGTGATGTTGCTGGACAGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 DB 781 GGTGGAATGACCAGATATGTTAGTATTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAATCTCAGATGGCCCTCTGGGCTATCATGTGCTGCTCTTTATTCTCTAAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACATATCAGCCTCAAGCCAAAGCTCTCTCTAGGATAGGAGCTAATTTGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 DB 961 CAGGACCCCTTTGGCAAGCAAGGTTACCACTTAGACAGGAGAGCAACTTTGAAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 DB 1021 GAAGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATATAAACCAGGAGAGATTGGT 1080
 QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGACTGCTGCTTTATACCATCGCATGCTTCCCTGGTAAAGGAGTGGCTGTATATCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
 DB 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAAGGAGCTAGGCTTCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 DB 1201 TCAAGGTTAAGAGTCAATTAATCCCAAGGACCTAGGCTTCTATGAATGGACT 1260
 QY 421 Met 421
 DB 1261 ATG 1263

RESULT 13

US-08-928-881-18
 ; Sequence 18, Application US/08928881
 ; Patent No. 6083725
 ; GENERAL INFORMATION:
 ; APPLICANT: Selden et al., Richard F.
 ; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A
 ; TITLE OF INVENTION: DEFICIENCY
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/928,881
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fraser, Janis K.
 ; REGISTRATION NUMBER: 34,819
 ; REFERENCE/DOCKET NUMBER: 07236/003001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/542-5070
 ; TELEFAX: 617/542-8906
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1343 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-928-881-18

Alignment Scores:
 Pred. No.: 8,84e-252 Length: 1343
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-08-928-881-18 (1-1343)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 36 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCCCTGGCC 95
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 96 CTCGTTCTCGGACATCTCTGGGCTAGAGCACTGGACATGGATGGCAAGGAGCCCT 155
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 156 ACCATGGGCTGGCTGCACCTGGAGCGCTTTCATGTGCAACCTTGGCTGCCAGGAGGCCA 215
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 216 GATTCCTGCATCAGTGAGAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 275
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 276 TGAAGGATGCAAGTTATGAGTACCTCTCATGTGATGACTGTGGATGGCTCCCAAGA 335
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheHisGlyIleArgGlnLeu 120
 Db 336 GATTCAGAAGGCAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 395
 QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
 Db 396 GCTAATATTGTTTACACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTGGAAATAA 455
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 Db 456 ACTTGGCAGGCTTCTCCCTGGAGTTTGGATAGTATGATGATGATGATGATGATGATGAT 515
 QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 516 GACTGGGGAGTAGACTGCTAAATTTGATGGTTGTTACTGTGACACTTTGGAAATTTG 575
 QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 Db 576 GCAGATGGTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGAGCATTTGTGTAC 635

QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
 Db 636 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 695
 QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 696 CAGTACTGCAATCACTGGCGAAATTTTCTGCACATTGATTTCTCTGGAAGATATAAG 755
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 756 AGTATCTTGGACTGGACATCTTTAAACAGGAGAGAATTGTGTATGTGTGTCGACAGGG 815
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 816 GGTGGATGACCCAGATATGTTAGTATTGGCACTTTGGCCTCAGCTGGAATCAGCAA 875
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 876 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTATTATTCATGCTTAATGACCTC 935
 QY 301 ArgHisIleSerProGlnAlaIleValLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 936 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGTAAATTCCTCAATCAAT 995
 QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 996 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGGAGAGCAACTTTGAAGTGG 1055
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1056 GAAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATCATATAAACCCGCGCAGGATTTGT 1115
 QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1116 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1175
 QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
 Db 1176 GCCTGCTTCATCACAGCTCTCTCTGTAAGAGGAGTGGTCTTCTATGAATGGACT 1235
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1236 TCAAGGTTAAGAAAGTCACATAAATCCACAGGCACTGTTTTGCTTTCAGCTAGAAAATACA 1295
 QY 421 Met 421
 Db 1296 ATG 1298

RESULT 14

US-09-543-921-18
 ; Sequence 18, Application US/09543921
 ; Patent No. 6395884
 ; GENERAL INFORMATION:
 ; APPLICANT: Seldin et al., Richard F.
 ; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A DEFICIENCY
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA: US/09/543,921
 ; FILING DATE: 06-Apr-2000
 ; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/928,881
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 07236/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1343 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-543-921-18
Alignment Scores:
Pred. No.: 8,846-252 Length: 1343
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-10-602-219-12 (1-421) x US-09-543-921-18 (1-1343)
QY 1 MetGlnLeuArgAspProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 36 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCGCTTCCTGGCC 95
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 96 CTCGHTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGAGGCT 155
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 156 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGTCCAGGAGAGGCA 215
QY 61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 216 GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 275
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 276 TGGAAAGGATGCAAGGTATGAGTACCTCTGCATGTATGACTGTGGATGGCTCCCAAGA 335
QY 101 AspSerGluGlyValGlnLeuAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 336 GATTCAGAAGGCAGACTTCAGGACAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 395
QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 396 GCTAATATGTTTCACAGCAAGAGACTGAAGCTAGGAGATTATGCAAGATGTTTGGAAATAA 455
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 456 ACCTGCGCAGGCTTCCTCGGAGTGTGGATATGATGATGATGATGATGATGATGATGATGAT 515
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 516 GACTGGGAGTAGACTGCTGAAAATTTGATGTTGTTACTGTGACAGATTGGAAAATTG 575
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 576 GCAGATGGTTATAAGCACATGCTCTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 635
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
DB 636 TCCTGTGAGTGGCCTCTTTATATGATGTCCTTTTCAAAAAGCCCAATTTATACAGAAATCCGA 695

QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 696 CAGTACTGCAATCACTGCGGAAATTTTCTGACATTTGATGATTTCTGGAAGATATAAG 755
QY 241 SerIleLeuAspTrpTrpSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 756 AGTATCTTGGACTGGACATCTTTTAAACCGAGAGAGATTTGTGATGTTGCTGGACAGGG 815
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 816 GGTGGAATGACCCAGATATGTTAGTATTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 875
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 876 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTCTGCTTAATGACCTC 935
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 936 CGACATCATGCCCTCAAGCCCAAGCTCTCTTCAGGATAGGAGTAAATTTGCCATCAAT 995
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 996 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1055
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1056 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTAGCTATGATATAAACCAGGAGATTGGT 1115
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1116 GGACCTGCTCTTATACCATGCGAGTTCTTCCCTGGGTAAGGAGTGGCTGTATCTCT 1175
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1176 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGTAGGTTCTATGAATGGACT 1235
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1236 TCAAGGTTTAAAGATGCATATAATCCACAGGCACTGTTTGTCTTCTCAGCTAGAAAATACA 1295
QY 421 Met 421
DB 1296 ATG 1298
RESULT 15
US-09-266-014-3
Sequence 3, Application US/09266014
Patent No. 6458574
GENERAL INFORMATION:
APPLICANT: Selden, Richard F
APPLICANT: Borowski, Marianne
APPLICANT: Kinoshita, Carol M
APPLICANT: Treco, Douglas A
APPLICANT: Williams, Melanie D
APPLICANT: Schuetz, Thomas J
APPLICANT: Daniel, Peter F
TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
FILE REFERENCE: FABRY DISEASE (18082-001)
CURRENT APPLICATION NUMBER: US/09/266,014
CURRENT FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: 60/026,041
PRIOR FILING DATE: 1996-09-13
PRIOR APPLICATION NUMBER: 08/928,881
PRIOR FILING DATE: 1997-09-12
PRIOR APPLICATION NUMBER: PCT US97/16603
PRIOR FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1343
TYPE: DNA
ORGANISM: Homo sapiens
US-09-266-014-3

Alignment Scores:

Pred. No.: 8,84e-252 Length: 1343
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-09-266-014-3 (1-1343)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 36 ATGCAGCTGAGGAACCCAGACTACATCTGGCTGGCGCTTGGCTTCCCTGGCC 95
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 96 CTCGTTTCTGGGACATCCCTGGGCTAGGCACTGGCAATGGATTGGCAAGGACGCT 155
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 156 ACCATGGCTGGCTGGGAGCGCTTCATGTGCACCTTGACTGCCAGGAAGGCCA 215
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 216 GATTCTTCATCAGTCAGAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 275
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 276 TGGAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 335
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 336 GATTTCAGAGGAGAGACTTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTGCCAGCTA 395
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 396 GCTAATATGTTACACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 455
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 456 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACTACGACATTGATGCCCGCAGACCTTTGCT 515
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 516 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGATTGGAAAATTGG 575
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 576 GCAGATGTTATAGACACATGCTTGGCCCTGAATAGGACTGGCAGAGCAATTGGTAC 635
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 636 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 695
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 696 CAGTACTGCAATCACATGGCGAAATTTGTCTGACATTGATGATTCCTGGAAAAGTATAAG 755
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 756 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGTGATGTTGCTGGACAGGG 815
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 816 GGTGGAAATGACCCAGATATGTTAGTGAATTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 875
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 876 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCCTCTTTATCATGTCTAATGACCTC 935
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
```

```
DB 936 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTGCCATCAAT 995
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 996 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1055
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1056 GNACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCCGACGAGATTGGT 1115
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1116 GGACCTCGCTCTTATATACCATCGCAGTTCCTTCCCTGGTAAAGGAGTGGCCTGTAACTCT 1175
QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1176 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1235
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1236 TCAAGTTTAAAGTACATTAATCCACAGGCACCTGTTTGTCTTCAGCTAGAAAATACA 1295
QY 421 Met 421
DB 1296 ATG 1298
```

Search completed: December 26, 2005, 13:52:44
Job time : 250 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 08:46:48 ; Search time 230 Seconds
(without alignments)
1291.424 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQRPELHLCALALRFLA.....RLRSHNPRTGVLLQLENTM 421

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	429	1 AGAL_HUMAN	P06280 homo sapien
2	2314	100.0	429	2 Q53Y83_HUMAN	Q53Y83 homo sapien
3	2313	100.0	429	2 Q53HF3_HUMAN	Q53HF3 homo sapien
4	2178	94.1	403	2 Q6LER7_HUMAN	Q6LER7 homo sapien
5	1811	78.3	419	1 AGAL_MOUSE	P51569 mus musculus
6	1811	78.3	421	2 Q8GZ6_MOUSE	Q8GZ6 mus musculus
7	1480.5	64.0	429	2 Q90WL7_FUGRU	Q90WL7 fugu rubrip
8	1267	54.8	377	2 Q4RTE7_TETNG	Q4RTE7 tetraodon n
9	1130	48.8	338	2 Q5XJ76_BRARE	Q5XJ76 brachydanio
10	1074.5	46.4	431	2 Q6GQ17_XENLA	Q6GQ17 xenopus lae
11	1056.5	45.7	400	2 Q6GR44_XENLA	Q6GR44 xenopus lae
12	1051	45.4	405	1 NAGAB_CHICK	Q90744 gallus gall
13	1050	45.4	412	2 Q7Q1V0_ANOGA	Q7Q1V0 anopheles g
14	1038	44.9	410	2 Q4STX0_TETNG	Q4STX0 tetraodon n
15	1031	44.6	413	2 Q9VL27_DROME	Q9VL27 drosophila
16	1031	44.6	427	2 Q8MYX3_DROME	Q8MYX3 drosophila
17	1007.5	43.5	419	2 Q7Q6H3_ANOGA	Q7Q6H3 anopheles g
18	1006	43.5	417	2 Q9V7N9_DROME	Q9V7N9 drosophila
19	990.5	42.8	415	1 NAGAB_MOUSE	Q9QWR8 mus musculus
20	982	42.4	415	1 NAGAB_RAT	Q66H12 rattus norv
21	967.5	41.8	411	1 NAGAB_BOVIN	Q58DH9 bos taurus
22	952.5	41.2	411	1 NAGAB_HUMAN	P17050 homo sapien
23	862	37.3	451	2 Q21801_CAEEL	Q21801 caenorhabdi
24	849	36.7	452	2 Q60LZ0_CABER	Q60LZ0 caenorhabdi
25	794	32.6	399	2 Q5DBS4_SCHUA	Q5DBS4 schistosoma
26	729	31.5	413	2 Q9FWV8_ORYSA	Q9FWV8 oryza sativ
27	716.5	31.0	434	2 Q9LYL2_ARATH	Q9LYL2 arabidopsis
28	716.5	31.0	437	2 Q8VXZ7_ARATH	Q8VXZ7 arabidopsis
29	712	30.8	378	1 AGAL_COPAR	Q42656 coffea arab
30	710	30.7	466	2 Q54EG4_DICDI	Q54EG4 dictyostell
31	710	30.7	478	2 Q5VVH6_CLOUO	Q5VVH6 clostridium

ALIGNMENTS				
RESULT 1				
ID	AGAL_HUMAN	STANDARD;	PRT;	429 AA.
AC	P06280;			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	10-MAY-2005 (Rel. 47, Last annotation update)			
DE	Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-			
DE	galactoside galactohydrolase) (Alpha-D-galactosidase A) (Agalsidase			
DE	alfa)			
GN	Name=GLA;			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;			
OC	Homo.			
OX	NCBI_TaxID=9606;			
RN	[1] _TaxID=9606;			
RP	NUCLEOTIDE SEQUENCE.			
RC	TISSUE=Lymphoblast;			
RC	MEDLINE=89263745; PubMed=2542896;			
RA	Kornreich R., Desnick R.J., Bishop D.F.;			
RT	"Nucleotide sequence of the human alpha-galactosidase A gene.;"			
RL	Nucleic Acids Res. 17:3301-3302(1989).			
RN	[2]			
RP	NUCLEOTIDE SEQUENCE.			
RC	TISSUE=Fibroblast;			
RC	MEDLINE=87246603; PubMed=3036505;			
RA	Tsuiji S., Martin B.M., Kaslow D.C., Migeon B.R., Choudary P.V.,			
RA	Stubblefield B.K., Mayor J.A., Murray G.J., Barranger J.A.,			
RA	Ginn S.E.;			
RT	"Signal sequence and DNA-mediated expression of human lysosomal alpha-			
RL	galactosidase A.;"			
RL	Eur. J. Biochem. 165:275-280(1987).			
RN	[3]			
RP	NUCLEOTIDE SEQUENCE.			
RC	MEDLINE=95352959; PubMed=7626884;			
RA	Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,			
RA	Gibbs R.A.;			
RT	"Sixty-nine kilobases of contiguous human genomic sequence containing			
RL	the alpha-galactosidase A and Bruton's tyrosine kinase loci.;"			
RL	Mamm. Genome 6:334-338(1995).			
RN	[4]			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RC	PubMed=15772651; DOI=10.1038/nature03440;			
RA	Ross M.T., Grafham D.V., Coffey A.J., Scherer S., McLay K., Muzny D.,			
RA	Platzer M., Howell G.R., Burrows C., Bird C.P., Frankish A.,			
RA	Lovell F.L., Howe K.L., Ashurst J.L., Fulton R.S., Sudbrak R., Wen G.,			
RA	Jones M.C., Huttles M.E., Deadman R., Carter N.P., Hunt S.E.,			
RA	Ramsey J., Whittaker A., Deadman R., Scott C.E., Searle S.,			
RA	Cree A., Gunaratne P., Havlak P., Hodgson A., Metzker M.L.,			
RA	Richards S., Scott G., Steffen D., Sodergren E., Wheeler D.A.,			
RA	Worley K.C., Ainscough R., Ambrose K.D., Ansari-Lari M.A., Aradhya S.,			
RA	Ashtwell R.I., Babbage A.K., Bagguley C.L., Ballabio A., Banerjee R.,			
RA	Barker G.E., Barlow K.F., Barrett I.P., Bates K.N., Beare D.M.,			

32	709	30.6	420	2	Q5DUH8_COPAR	Q5DUH8 coffea arab
33	698.5	30.2	425	2	Q7XIV4_ORYSA	Q7XIV4 oryza sativ
34	697.5	30.1	431	2	Q5XTZ3_SALMI	Q5XTZ3 salvia milt
35	695.5	30.1	380	2	Q9FUR8_LYCES	Q9FUR8 lycopersico
36	695.5	30.1	409	2	Q9SP05_LYCES	Q9SP05 lycopersico
37	694	30.0	378	2	Q5DUH7_COFCA	Q5DUH7 coffea cane
38	692	29.9	428	2	Q84VQ7_HELAN	Q84VQ7 helianthus
39	687.5	29.7	417	1	AGAL_ORYSA	Q9FTX4 oryza sativ
40	687.5	29.7	422	2	Q39811_SOYBN	Q39811 glycine max
41	686	29.6	385	2	Q54AX7_DICDI	Q54AX7 dictyostell
42	684.5	29.6	396	2	Q8RX86_ARATH	Q8RX86 arabidopsis
43	682	29.5	425	2	Q41100_PHAVU	Q41100 phaseolus v
44	681.5	29.5	411	1	AGAL_CYATE	P14749 cyanopsidis t
45	681	29.4	405	2	Q5ZP79_PEA	Q5ZP79 pisum sativ

RA Beasley H., Beasley O., Beck A., Bethel G., Blechschmidt K., Brady N.,
 RA Bray-Allen S., Bridgeman A.M., Brown A.J., Brown M.J., Bonnin D.,
 RA Bruford E.A., Buhay C., Burch P., Burford D., Burgess J., Burrill W.,
 RA Burton J., Bye J.M., Carder C., Carrel L., Chako J., Chapman J.C.,
 RA Chavez D., Chen E., Chen G., Chen Y., Chen Z., Chinault C.,
 RA Ciccodicola A., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
 RA Clerc-Blankenburg K., Clifford K., Cobley V., Cole C.G., Conquer J.S.,
 RA Corby N., Connor R.E., David R., Davies J., Davis C., Davis J.,
 RA Delgado O., Deshazo D., Dhali P., Ding Y., Dinh H., Dodsworth S.,
 RA Draper H., Dugan-Rocha S., Dunham A., Dunn M., Durbin K.J., Dutta I.,
 RA Eades T., Ellwood M., Emery-Cohen A., Erington H., Evans K.L.,
 RA Faulkner L., Francis F., Frankland J., Fraser A.E., Galgoczy P.,
 RA Gilbert J., Gill R., Gloeckner G., Gregory S.G., Gribble S.,
 RA Griffiths C., Grocock R., Gu Y., Gwilliam R., Hamilton C., Hart E.A.,
 RA Hawes A., Heath P.D., Heitmann K., Hennig S., Hernandez J.,
 RA Hinzmann B., Ho S., Hoffs M., Howden P.J., Huckle E.J., Hume J.,
 RA Hunt P.J., Hunt A.R., Isherwood J., Jacob L., Johnson D., Jones S.,
 RA de Jong P.J., Joseph S.S., Keenan S., Kelly S., Kershaw J.K., Khan Z.,
 RA Kioschis P., Klages S., Knights A.J., Koslura A., Kovar-Smith C.,
 RA Laird G.K., Langford C., Lawlor S., Leverisha M., Lewis L., Liu W.,
 RA Lloyd C., Lloyd D.M., Loulsegod H., Loveland J.E., Lovell J.D.,
 RA Lozado R., Lu J., Lyne R., Ma J., Maheshwari M., Matthews L.H.,
 RA McDowall J., Mettaren S., McMurray A., Meidl P., Meitinger T.,
 RA Milne S., Miner G., Mistry S.L., Morgan M., Morris S., Mueller I.,
 RA Mullikin J.C., Nguyen N., Nordsiek G., Nyakatura G., O'dell C.N.,
 RA Okuwonu G., Palmer S., Pandian R., Parker D., Parrish J.,
 RA Perez L., Porter K.M., Ramsey Y., Reichwald K., Rhodes S.,
 RA Rädler K.A., Schlesinger D., Schueler M.G., Sehra H.K.,
 RA Shaw-Smith C., Shen H., Sheridan E.M., Showkeen R., Skuce C.D.,
 RA Smith M.B., Sotharan E.C., Steingrubner H.E., Steward C.A., Storey R.,
 RA Swann R.M., Swarbreck D., Taber P.E., Taudien S., Taylor T.,
 RA Teague B., Thomas K., Thorpe A., Timms K., Tracey A., Trevanion S.,
 RA Tromans A.C., d'Uro M., Verdusco D., Villasana D., Waldron L.,
 RA Wall M., Wang Q., Warren J., Warry G.L., Wei X., West A.,
 RA Whitehead S.B., Whiteley M.N., Wilkinson J.E., Willey D.L.,
 RA Williams G., Williams L., Williamson A., Williamson H., Wilming L.,
 RA Woodmansey R.L., Wray P.W., Yen J., Zhang J., Zhou J., Zoghbi H.,
 RA Zorilla S., Buck D., Reinhardt R., Poustka A., Rosenthal A.,
 RA Lehrach H., Meindl A., Minx P.J., Hillier L.W., Willard H.F.,
 RA Wilson R.K., Waterston R.H., Rice C.M., Vaudin M., Coulson A.,
 RA Nelson D.B., Weinstock G., Sulston J.E., Durbin R., Hubbard T.,
 RA Gibbs R.A., Beck S., Rogers J., Bentley D.R.;
 RA "The DNA sequence of the human X chromosome.";
 RA Nature 434:325-337(2005).
 [5]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Uterus;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie H.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smallus D.B.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [6]
 RP NUCLEOTIDE SEQUENCE OF 31-429, AND PARTIAL PROTEIN SEQUENCE.
 RC TISSUE=Lung;
 RX MEDLINE=86259694; PubMed=3014515;
 RA Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
 RA Desnick R.J.;
 RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
 RT encoding the mature enzyme.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863(1986).
 [7]
 RP NUCLEOTIDE SEQUENCE OF 1-64.
 RX MEDLINE=88112869; PubMed=2892762; DOI=10.1016/0378-1119(87)90374-X;
 RA Quinn M., Hantzopoulos P., Fidanza V., Calhoun D.H.;
 RT "A genomic clone containing the promoter for the gene encoding the
 RT human lysosomal enzyme, alpha-galactosidase A.";
 RL Gene 58:177-188(1987).
 [8]
 RP NUCLEOTIDE SEQUENCE OF 1-64.
 RX MEDLINE=88234528; PubMed=2836863;
 RA Bishop D.F., Kornreich R., Desnick R.J.;
 RT "Structural organization of the human alpha-galactosidase A gene:
 RT further evidence for the absence of a 3' untranslated region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:3903-3907(1988).
 [9]
 RP RNA EDITING OF POSITION 396.
 RX MEDLINE=95380278; PubMed=7503918;
 RA Novo F.J., Kruszewski A., McDermot K.D., Goldspink G., Gorecki D.C.;
 RT "Editing of human alpha-galactosidase RNA resulting in a pyrimidine to
 RT purine conversion.";
 RL Nucleic Acids Res. 23:2636-2640(1995).
 [10]
 RP X-RAY CRYSTALLOGRAPHY (3.45 ANGSTROMS) OF 32-422 IN COMPLEX WITH
 RP PRODUCT, HOMODIMERIZATION, AND N-GLYCOSYLATION.
 RX PubMed=15003450; DOI=10.1016/j.jmb.2004.01.035;
 RA Garman S.C., Garboczi D.N.;
 RT "The molecular defect leading to Fabry disease: structure of human
 RT alpha-galactosidase.";
 RL J. Mol. Biol. 337:319-335(2004).
 [11]
 RP REVIEW ON FD VARIANTS.
 RX MEDLINE=94258158; PubMed=7911050;
 RA Eng C.M., Desnick R.J.;
 RT "Molecular basis of Fabry disease: mutations and polymorphisms in the
 RT human alpha-galactosidase A gene.";
 RL Hum. Mutat. 3:103-111(1994).
 [12]
 RP VARIANT FD SER-40.
 RX MEDLINE=90092580; PubMed=2152885; DOI=10.1016/0014-5793(90)80046-L;
 RA Koide T., Ishiura M., Iwai K., Inoue M., Kaneda Y., Okada Y.,
 RA Uchida T.;
 RT "A case of Fabry's disease in a patient with no alpha-galactosidase A
 RT activity caused by a single amino acid substitution of Pro-40 by
 RT Ser.";
 RL FEBS Lett. 259:353-356(1990).
 [13]
 RP VARIANT FD VAL-296.
 RX MEDLINE=91101674; PubMed=1846223;
 RA von Scheidt W., Eng C.M., Fitzmaurice T.F., Erdmann E., Hubner G.,
 RA Olsen E.G.J., Christomanou H., Kandolf R., Bishop D.F., Desnick R.J.;
 RT "An atypical variant of Fabry's disease with manifestations confined
 RT to the myocardium.";
 RL N. Engl. J. Med. 324:395-399(1991).
 [14]
 RP VARIANT FD GLN-301.
 RX MEDLINE=91027271; PubMed=2171331;
 RA Sakuraba H., Oshima A., Fukuhara Y., Shimamoto M., Nagao Y.,
 RA Bishop D.F., Desnick R.J., Suzuki Y.;
 RT "Identification of point mutations in the alpha-galactosidase A gene
 RT in classical and atypical hemizygotes with Fabry disease.";
 RL Am. J. Hum. Genet. 47:784-789(1990).
 [15]
 RP VARIANT FD TRP-356.
 RX MEDLINE=89198098; PubMed=2539398;
 RA Bernstein H.S., Bishop D.F., Astrin K.H., Kornreich R., Eng C.M.,
 RA Sakuraba H., Desnick R.J.;
 RT "Fabry disease: six gene rearrangements and an exonic point mutation
 RT in the alpha-galactosidase gene."

```

RL J. Clin. Invest. 83:1390-1399(1989).
RN [16]

Query Match      100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.8e-187;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPDLHGICALALRFLALVSWDIPGARALDNGGLARTPTMGWLHWFPMCNLDQCEEP 60
DB 1 MOLRNPDLHGICALALRFLALVSWDIPGARALDNGGLARTPTMGWLHWFPMCNLDQCEEP 60

QY 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDWCWAPQDSEGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDWCWAPQDSEGRLOADPQRPFGIRQL 120

QY 121 ANYVHSGKGLGIYADVGNKTCAGFPFGSGFYDYDDAQTADWGVLLKFDGVCYCDSENL 180
DB 121 ANYVHSGKGLGIYADVGNKTCAGFPFGSGFYDYDDAQTADWGVLLKFDGVCYCDSENL 180

QY 181 ADGKMSLALNRTGRSIVYSCWPLYMPPFKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKMSLALNRTGRSIVYSCWPLYMPPFKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

QY 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSWNQVOTOMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSWNQVOTOMALWAIMAAPLFMSNDL 300

QY 301 RHISPOAKALLQDKVIAINQDPLGKQYQOLRGDNFVWERPLSGLAWAVAMINRQEIIG 360
DB 301 RHISPOAKALLQDKVIAINQDPLGKQYQOLRGDNFVWERPLSGLAWAVAMINRQEIIG 360

QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420

QY 421 M 421
DB 421 M 421

RESULT 2
Q53Y83 HUMAN
ID Q53Y83_HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53Y83;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Galactosidase, alpha.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Kounidinya M., Raphael J., Moreira D., Kelley T., Labaer J., Lin Y.,
RA Phelan M., Farmer A.;
RT "Cloning of human full-length cDNAs in BD Creator(TM) System Donor
RT vector."
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BF006864; AAP35510.1; -; mRNA.
SQ SEQUENCE 429 AA; 48766 MW; 613F8BF21B107D7B CRC64;

Query Match      100.0%; Score 2314; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.8e-187;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPDLHGICALALRFLALVSWDIPGARALDNGGLARTPTMGWLHWFPMCNLDQCEEP 60
DB 1 MOLRNPDLHGICALALRFLALVSWDIPGARALDNGGLARTPTMGWLHWFPMCNLDQCEEP 60

QY 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDWCWAPQDSEGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDWCWAPQDSEGRLOADPQRPFGIRQL 120

QY 121 ANYVHSGKGLGIYADVGNKTCAGFPFGSGFYDYDDAQTADWGVLLKFDGVCYCDSENL 180
DB 121 ANYVHSGKGLGIYADVGNKTCAGFPFGSGFYDYDDAQTADWGVLLKFDGVCYCDSENL 180

QY 181 ADGKMSLALNRTGRSIVYSCWPLYMPPFKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKMSLALNRTGRSIVYSCWPLYMPPFKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

QY 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSWNQVOTOMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSWNQVOTOMALWAIMAAPLFMSNDL 300

QY 301 RHISPOAKALLQDKVIAINQDPLGKQYQOLRGDNFVWERPLSGLAWAVAMINRQEIIG 360
DB 301 RHISPOAKALLQDKVIAINQDPLGKQYQOLRGDNFVWERPLSGLAWAVAMINRQEIIG 360

QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420

QY 421 M 421
DB 421 M 421

RESULT 3
Q53HF3 HUMAN
ID Q53HF3_HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53HF3;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Galactosidase, alpha variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Maruyama K., Sugano S.;
RT "Oligo-capping: a simple method to replace the cap structure of
RT eucaryotic mRNAs with oligoribonucleotides."
RL Gene 138:171-174 (1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Yoshitomo K., Maruyama K., Sugano A., Sugano S.;
RT "Construction and characterization of a full length-enriched and a 5'-
RT end-enriched cDNA library."
RL Gene 200:149-156(1997).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
RA Tanaka A., Yokoyama S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK222627; BAD96347.1; -; mRNA.
FT NON_TER 1
SQ SEQUENCE 429 AA; 48752 MW; AFA828F5B22EED76 CRC64;

Query Match      100.0%; Score 2313; DB 2; Length 429;
Best Local Similarity 99.8%; Pred. No. 3.4e-187;
Matches 420; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPDLHGICALALRFLALVSWDIPGARALDNGGLARTPTMGWLHWFPMCNLDQCEEP 60
DB 1 MOLRNPDLHGICALALRFLALVSWDIPGARALDNGGLARTPTMGWLHWFPMCNLDQCEEP 60

```


QY	61	DSICSEKLFMEAEMLVSEGWKADAGYEYLCIDDWMAPOQDSEGRLOADPQPFPHGIROL	120	DB	1	PGARALDNGLARTPTMGWLHWRFCNLDCCQEPDSCISEKLFMEAEMLVSEGWKADAGY	60
DB	61	DSICSEKLFMEAEMLVSEGWKADAGYEYLCIDDWMAPOQDSEGRLOADPQPFPHGIROL	120	QY	87	EYLCIDDDCWMAPOQDSEGRLOADPQPFPHGIROLANYVHSGKLGKGIYADVGNKTCAGFP	146
QY	121	ANYVHSGKLGKGIYADVGNKTCAGPPGSGYGYDIDAQTFADGWVLLKFDGVCDSLENL	180	DB	61	EYLCIDDDCWMAPOQDSEGRLOADPQPFPHGIROLANYVHSGKLGKGIYADVGNKTCAGFP	120
DB	121	ANYVHSGKLGKGIYADVGNKTCAGPPGSGYGYDIDAQTFADGWVLLKFDGVCDSLENL	180	QY	147	GSGFYGYDIDAQTFADGWVLLKFDGVCDSLENLADGKHSALNRTGRSIVYSCWPL	206
QY	181	ADGKHSALNRTGRSIVYSCWPLMPPFKPNTYETROYCNHWRNFADIDDSWKS	240	DB	121	GSGFYGYDIDAQTFADGWVLLKFDGVCDSLENLADGKHSALNRTGRSIVYSCWPL	180
DB	181	ADGKHSALNRTGRSIVYSCWPLMPPFKPNTYETROYCNHWRNFADIDDSWKS	240	QY	207	YMWPKQKNYETROYCNHWRNFADIDDSWKSIGILDTWTSFNQERIVDVVAGPGGWNDD	266
QY	241	SILDWTSFNQERIVDVVAGPGGWNDDPKMLVIGNFGLSWNQVQTMALWAIMAAPLFMSNDL	300	DB	181	YMWPKQKNYETROYCNHWRNFADIDDSWKSIGILDTWTSFNQERIVDVVAGPGGWNDD	240
DB	241	SILDWTSFNQERIVDVVAGPGGWNDDPKMLVIGNFGLSWNQVQTMALWAIMAAPLFMSNDL	300	QY	267	MLVIGNFGLSWNQVQTMALWAIMAAPLFMSNDLRHSPOAKALLQDKDVIAINQDPLGK	326
QY	301	RHISPOAKALLQDKDVIAINQDPLGKQVQLRQGDNFVEWVERPLSGLAWAVAMINROEIG	360	DB	241	MLVIGNFGLSWNQVQTMALWAIMAAPLFMSNDLRHSPOAKALLQDKDVIAINQDPLGK	300
DB	301	RHISPOAKALLQDKDVIAINQDPLGKQVQLRQGDNFVEWVERPLSGLAWAVAMINROEIG	360	QY	327	QGYQLRQGDNFVEWVERPLSGLAWAVAMINROEIGGPRSRYSYTIASVSLGKGVACNPACFITQ	386
QY	361	GPRSYTIASVSLGKGVACNPACFITQLLPVKKRLGFEWTSRLRSHINPTGTVLLQLENT	420	DB	301	QGYQLRQGDNFVEWVERPLSGLAWAVAMINROEIGGPRSRYSYTIASVSLGKGVACNPACFITQ	360
DB	361	GPRSYTIASVSLGKGVACNPACFITQLLPVKKRLGFEWTSRLRSHINPTGTVLLQLENT	420	QY	387	LLPVKKRLGFEWTSRLRSHINPTGTVLLQLENTM	421
QY	421	M 421		DB	361	LLPVKKRLGFEWTSRLRSHINPTGTVLLQLENTM	395
DB	421	M 421		RESULT 5			
Q6LER7_HUMAN							
ID	Q6LER7_HUMAN	PRELIMINARY;	PRT;	403	AA.		
AC	Q6LER7	2004 (TrEMBLrel. 27, Created)					
DT	05-JUN-2004	(TrEMBLrel. 27, Last sequence update)					
DT	05-JUN-2004	(TrEMBLrel. 27, Last annotation update)					
DE	Alpha-galactosidase A (EC 3.2.1.22) (Fragment).						
GN	Name=alpha-Gala;						
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;						
OC	Homo.						
OX	NCBI_TaxID=9606;						
RN	[1]						
RP	NUCLEOTIDE SEQUENCE.						
RC	TISSUE=Lung;						
RX	MEDLINE=86259694; PubMed=3014515;						
RA	Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,						
RA	Denick R.J.;						
RT	"Human alpha-galactosidase A: nucleotide sequence of a cDNA clone						
RT	encoding the mature enzyme."						
RL	Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863 (1986).						
DR	EMBL; D00039; BAA34059.1; -; mRNA.						
DR	SMR; Q6LER7; 6-396.						
DR	GO; GO:0004557; F:alpha-galactosidase activity; IEA.						
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.						
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.						
DR	InterPro; IPR002241; Glyco_hydro_27.						
DR	InterPro; IPR000111; Glyco_hydro_GHD.						
DR	Pfam; PF02065; Melibiase; 1.						
DR	PRINTS; PR00740; GLHYDRLASE27.						
DR	ProDom; PD002572; Glyco_hydro_GHD; 1.						
DR	PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.						
KW	Glycosidase; Hydrolase.						
FT	CHAIN	6	403	alpha-galactosidase A subunit peptide.			
FT	NON_TER	1	1				
SQ	SEQUENCE	403	AA;	45804	NW;	2FE193205BEB8D1A	CRC64;
Query Match							
Best Local Similarity				94.1%	Score	2178;	DB 2; Length 403;
Matches 395; Conservative				100.0%	Pred. No.	8.6e-176;	Indels 0; Gaps 0;
				0;	Mismatches	0;	
QY	27	PGARALDNGLARTPTMGWLHWRFCNLDCCQEPDSCISEKLFMEAEMLVSEGWKADAGY	86				

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedon T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahay J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield A., Schlein J.E., Jones S.J.M., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-
CC galactose residues in alpha-D-galactosides, including galactose
CC oligosaccharides, galactomannans and galactohydrolase.
CC -I- SUBUNIT: Homodimer.
CC -I- SUBCELLULAR LOCATION: Lysosomal.
CC -I- SIMILARITY: Belongs to the glycosyl hydrolase 27 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC DR EMBL: U34071; AAA96749.1; -; mRNA.
CC DR EMBL: L46651; AAA74453.1; -; Genomic DNA.
CC DR EMBL: U58105; ABA7244.1; -; Genomic DNA.
CC DR EMBL: U50716; AAC52584.1; -; mRNA.
CC DR EMBL: U50715; AAC52583.1; -; Genomic DNA.
CC DR EMBL: BC009021; AAH09021.1; -; mRNA.
CC DR PIR: JC4522; JC4522.
CC DR SMR: P51569; 32-418.
CC DR Ensembl: ENSMUSG00000031266; Mus musculus.
CC DR MGI: MG1:1347344; Gla.
CC DR GO: GO:0005615; C:extracellular space; TAS.
CC DR InterPro: IPR002241; Glyco_hydro_27.
CC DR InterPro: IPR000111; Glyco_hydro_GHD.
CC DR Pfam: PF02065; Melibiase; 1.
CC DR PRINTS: PR00740; GLHYDRLASE27.
CC DR ProDom: PD002572; Glyco_hydro_GHD; 1.
CC DR PROSITE: PS00512; ALPHA_GALACTOSIDASE; 1.
CC KW Glycoprotein; Glycosidase; Hydrolase; Lysosome; Signal.
CC FT SIGNAL 1 31 By similarity.
CC FT CHAIN 32 419 Alpha-galactosidase A.
CC FT ACT_SITE 170 170 Nucleophile (By similarity).
CC FT ACT_SITE 231 231 Proton donor (By similarity).
CC FT SITE 203 207 Substrate binding (By similarity).
CC FT CARBOHYD 139 139 N-linked (GlcNAc...) (By similarity).
CC FT CARBOHYD 192 192 N-linked (GlcNAc...) (By similarity).
CC FT CARBOHYD 215 215 N-linked (GlcNAc...) (By similarity).
CC FT DISULFID 52 94 By similarity.
CC FT DISULFID 56 63 By similarity.
CC FT DISULFID 142 172 By similarity.
CC FT DISULFID 202 223 By similarity.
CC FT DISULFID 378 382 By similarity.
CC FT SEQUENCE 419 AA; 47643 MW; BD5E5A99AC113613 CRC64;
Query Match 78.3%; Score 1811; DB 1; Length 419;
Best Local Similarity 78.2%; Pred. No. 1.1e-144;
Matches 327; Conservative 41; Mismatches 50; Indels 0; Gaps 0;
QY 1 MQLRNPDLHLCALRFLAVSWDIPGARLDNGLARTPTWGLWHLWRFMNCNDQCEP 60
DB 1 MKLLSRDRLVCELALCPALVFWISILGVRALDNGLARTPTWGLWHLWRFMNCNDQCEP 60

QY 61 DSCISKLFPMWAEALMVSGWKDAGYEYLCIDCWMAQORDSEGRLOADPQRPFGHQR 120
DB 61 DACISQLPQMAELMVSGWRDAGYDYLICIDCWMAQORDSEGRLOADPQRPFGHQR 120
QY 121 ANYVHSGKLGIGYADVGNKTCAGPFGSGGYDYDIDQATFADWGVLLKPDGCYCDLSLNL 180
DB 121 ANYVHSGKLGIGYADVGNKTCAGPFGSGGYDYDIDQATFADWGVLLKPDGCYCDLSVSL 180
QY 181 ADGKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ENGKYMALALNRTGRSIVYSCWPLYLRFPHKPNYTDIQYCNHWRNFDDVYDSWESIK 240
QY 241 SILDWTSFNQERIVDVAGPGGWNDDPMLVIGNFLGSWNOQVTOALWAIMAAFLFMSNDL 300
DB 241 NILSWTVVYKEIVEVAGPGGWNDDPMLVIGNFLGSWNOQVTOALWAIMAAFLFMSNDL 300
QY 301 RHISPOKALLQDKQVIAINQDPLGKQGYQLRQGDNFVWERPLSLGLAWAVAMINQEIG 360
DB 301 ROISSQAKALLQDKQVIAINQDPLGKQGYCFRKNHIEVWERPLSLGLAWAVAVRNQEIG 360
QY 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLE 418
DB 361 GPCPTYIQISSLGRGLACNPGCIIITQLPEKVLHGFYEWTLTKTRVNSGTVLFRLE 418
RESULT 6
Q8BGZ6_MOUSE PRELIMINARY; PRT; 421 AA.
AC Q8BGZ6_MOUSE
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A430057F16 product:galactosidase, alpha, full insert
DE sequence (Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
DE full-length enriched library, clone:E330039P08 product:galactosidase,
DE alpha, full insert sequence).
GN Name=Gla;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RX Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RX Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RX Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RX Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RX Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RX Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RX Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RX Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RX Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RX Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RX Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RX Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RX Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RX Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RX Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-P.,
RX Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RX Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RX Hayashizaki Y.;

Wed Dec 28 11:15:36 2005

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RA Director MGC Project;
 RL Submitted (OCT-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC083209; AAH83209.1; -; mRNA.
 DR ZFIN; ZDB-GENE-041010-207; zgc:101584.
 DR GO; GO:0004553; F:hydroxylase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR002241; Glyco_hydro_27.
 DR InterPro; IPR000111; Glyco_hydro_GHD.
 DR PRINTS; PR00740; GLHYDRLASE27.
 DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
 DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
 FT NON TER 1
 SQ SEQUENCE 338 AA; 37846 MW; A3FC5C461B7C9649 CRC64;

Query Match 48.8%; Score 1130; DB 2; Length 338;
 Best Local Similarity 60.3%; Pred. No. 4.1e-87;
 Matches 210; Conservative 47; Mismatches 75; Indels 16; Gaps 2;

QY 68 LFMEMAEIYVSEGVKADAGYEYLCIDDCWMAPOQDSEGRLOADPQRPFGHRIQLANYVHSK 127
 DB 2 LDHSLCALMTA-GLRNKG-----THRGACRQTPKGFVPVASKLADYVHSK 45
 QY 128 GLKLGIVADGNKTCAGPFGSGFYDIDAQTFADWGVLDLKFDCYCDSENLDGKYHM 187
 DB 46 GLKLGIVADGVTKTCAGPFGSLGYDIDAKTFADWGVLDLKFDCGFMDFHQLGEGYINM 105
 QY 188 SLALNRTGRSIVSCPEPLVMPQKPNYTEIROYCNHWNFAIDDDSWKSIXSLDWT 247
 DB 106 SSALNQTGRSIVSCPEPLVMPQKPNYTEIROYCNHWNFAIDDDSWKSIXSLDWT 165
 QY 248 FQNRIVDVAGPGGWNPDMLVINGFLSNQOQVTOALMAIAPLFMSNDLRHISPOA 307
 DB 166 EKQIVVPVAGPGGWNPDMLVINGFLSNQOQVTOALMAIAPLFMSNDLRHISPOA 225
 QY 308 KALLQDQXDVIAINQDPLGKQYQLRQGNFVWERPLSLGAWAVAMINRQETGG 367
 DB 226 KELLQNKQIIAINDPLGKQYQLRQGNFVWERPLSLGAWAVAMINRQETGG 285
 QY 368 AVASIGKGVACNPACFIQLLPVKKLGFYEWTSRLRSHINPTGTVLL 415
 DB 286 SVAIMPSKLCNPKCNVTQILPTYKEMGVQNLSEVVVQVNPVTGTTLL 333

RESULT 10

Q6GQ17 XENLA
 ID Q6GQ17_XENLA PRELIMINARY; PRT; 431 AA.
 AC Q6GQ17;
 DT 05-JUL-2004 (TREMBLrel. 27, Created)
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
 DE LOC443592 protein (fragment).
 GN Name=LOC443592; (African clawed frog).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 Dev. Dyn. 225:384-391 (2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC072931; AAH72931.1; -; mRNA.
 DR GO; GO:0004553; F:hydroxylase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR002241; Glyco_hydro_27.
 DR InterPro; IPR000111; Glyco_hydro_GHD.
 DR PRINTS; PR00740; GLHYDRLASE27.
 DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
 DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
 FT NON TER 1
 SQ SEQUENCE 431 AA; 48787 MW; 4C578DB433983471 CRC64;

Query Match 46.4%; Score 1074.5; DB 2; Length 431;
 Best Local Similarity 49.3%; Pred. No. 2.8e-82;
 Matches 209; Conservative 62; Mismatches 126; Indels 27; Gaps 6;

QY 4 RNPELHGCALALRFLALVSWDIPGALDNGLARTTMCGLHWRFCNLDQCEEPDSC 63
 DB 18 RGYKHLTSLCLLTLMGLCW-----CLDNLGLVTPPMGMWMTQRYRCNIDCKSDPNC 71
 QY 64 ISEKLFMEAEIYVSEGVKADAGYEYLCIDDCWMAPOQDSEGRLOADPQRPFGHRIQLANY 123
 DB 72 ISENLIKIMADEMADSGWRDLGYVVICIDDCWSQKQSDNGRLQDPERPFGMKALADY 131
 QY 124 VHSKGLKLGIVADGNKTCAGPFG-SFGYDIDAQTFADWGVLDLKFDCYCDSENLD 182
 DB 132 VHAKGLKLGIVSDMGNTYTCGYPCTLTIDTAETFASWEVDMLKFDCGYSNSTEK-AL 190
 QY 183 GYKMSLALNRTGRSIVSCPEPLVMPQKPNYTEIROYCNHWNFAIDDDSWKSIXS 241
 DB 191 GYPKNSEALNATGRPILYSCSWPAYEGGLPPKVNVTQLGSTCNMWRNYDDDDSDWSVLD 250
 QY 242 ILDTWTSFQNRIVDVAGPGGWNPDMLVINGFLSNQOQVTOALMAIAPLFMSNDLR 301
 DB 251 IIEWAKNQDVLQPAAGRWNDPDLITGDFGLSYEQSKQLAIWALAPLMSNDLR 310
 QY 302 HISPOAKALQDQXDVIAINQDPLGKQYQLRQGNFVWERPLSLGAWAVAMINRQETGG 361
 DB 311 AISQADKDLLQNRLLIYINQSLGKQSGSLISRIENLGVKRELINGQYALAVLNKGTDL 370
 QY 362 PRSYTIAVASIGKGVACNPACFIQLLP-----VKKLGFYEWTSRLRSHINPTGT 413
 DB 371 PRPYSTNLGLLN-----VTQCTDGYKLYNVFKEYGLGMFTSTPIIRVNPTGT 420


```
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRLASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW 3D-structure; Direct protein sequencing; Glycoprotein; Glycosidase;
KW Hydrolase; Lysosome.
FT CHAIN 1 405 Alpha-N-acetylgalactosaminidase.
FT ACT_SITE 140 140 Nucleophile (By similarity).
FT ACT_SITE 201 201 Proton donor (By similarity).
FT CARBOHYD 161 161 N-linked (GlcNAc... ).
FT CARBOHYD 185 185 N-linked (GlcNAc... ).
FT CARBOHYD 369 369 N-linked (GlcNAc... ).
FT DISULFID 21 63
FT DISULFID 25 32
FT DISULFID 111 142
FT DISULFID 171 193
SQ SEQUENCE 405 AA; 45615 MW; E1EC0061739C305C CRC64;

Query Match 45.4%; Score 1051; DB 1; Length 405;
Best Local Similarity 53.0%; Pred. No. 2.5e-80;
Matches 206; Conservative 58; Mismatches 117; Indels 8; Gaps 6;

QY 32 LONGLARTPTGMLHWRFCNLDCEBPDSICSEKLFMEAEMLVSEGWKADAGEYLICI 91
DB 1 LENGLARTPPMGWLAWERFCNVCNCRDPRQCISEMLFMEMADRIAEDGWRELGYKYINI 60

QY 92 DDCWMAQRDSEGRLOADPQRPFGHIGRLQANYVHSGKLGIVADVGNKTCAGPFG-SFG 150
DB 61 DDCWAAKORDAEGRLVDPFPRFGIRKALADYVHARGLKLDIYDGLRLTCGGPGVTILD 120

QY 151 YYDIDAQTADGWGVDLLKFDGCGYCDLSLENLADGYKMSLALNRTGRSIVYSCWPLYMWP 210
DB 121 RVEQDAQTFAEMGVNMLKLDGYSYGKEQ-AQGYPPQMARALNSTGRPIVYSCSWPAFOGG 179

QY 211 F-QKPNTEIRQYCNHNRNFADIDDSWKSILDTWTSFNOERIVDVAGFGWNPDMIV 269
DB 180 LPPKVNVTLLGEICNLWRNVDIQQSDWSVLISVDWFTNQDVLQVFPAGEFGHWNPDMLI 239

QY 270 IGNGLSNQQVOTQMALWAIAPLMSNDRHISPOAKALLQDKVDVIAINQDPLGKQGY 329
DB 240 IGNGLSYEQSRQSMALWTMAAPLMSNDRHISPOAKALLQDKVDVIAINQDPLGKQGR 299

QY 330 Q-LRQGNFYEYWRPLSGLAWAMINRQETGGPSRYTIVASLGKGVACNPACFIITQLL 388
DB 300 RIIEGSHIEVFLPLSQASALVFFSRR-TDMFFRYTTSLSKLGFPWG---AAVEQDV 355

QY 389 PVKRLGFEYWTSLRSHINPTGVLLQL 417
DB 356 YSGKIISGLTKGDNFTVIINPNSGVVMWYL 384
```

```
RESULT 13
Q7Q1V0 ANOGA
ID Q7Q1V0 ANOGA PRELIMINARY; PRT; 412 AA.
AC Q7Q1V0;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000020847 (fragment).
GN ORFNames=ENSANG0000018359;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```

```
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008980; EAA14548.2; -; Genomic DNA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRLASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
FT NON_TER 1 1
FT NON_TER 412 412
SQ SEQUENCE 412 AA; 46517 MW; 4C91C8FB9EA6C88F CRC64;

Query Match 45.4%; Score 1050; DB 2; Length 412;
Best Local Similarity 49.0%; Pred. No. 3.1e-80;
Matches 193; Conservative 72; Mismatches 119; Indels 10; Gaps 4;

QY 31 ALDNGLARTPTGMLHWRFCNLDCEBPDSICSEKLFMEAEMLVSEGWKADAGEYLIC 90
DB 17 SLENGLARTPPMGWLAWERFCNVCNCRDPRQCISEMLFMEMADRIAEDGWRELGYKYINI 76

QY 91 IDCWMAQRDSEGRLOADPQRPFGHIGRLQANYVHSGKLGIVADVGNKTCAGPFGSFG 150
DB 77 VDCWLEKSRGPRGELVADRPRFSGWKALANVYHAKLGKFGIVDGYNTCAGYPGILG 136

QY 151 YYDIDAQTADGWGVDLLKFDGCGYCDLSLENLADGYKMSLALNRTGRSIVYSCWPLY-MW 209
DB 137 FSANDAQAQFASWDVYVLDGCGYSLPID-MDHGYPEFGRLNATGRPMIYSCSWPVYQIY 195

QY 210 PFQKPNTEIRQYCNHNRNFADIDDSWKSILDTWTSFNOERIVDVAGFGWNPDMIV 269
DB 196 AGWNPYSSIIQCNLWRNVDIQQSDWSVLISVDWFTNQDVLQVFPAGEFGHWNPDMLI 255

QY 270 IGNGLSNQQVOTQMALWAIAPLMSNDRHISPOAKALLQDKVDVIAINQDPLGKQGY 329
DB 256 IGNGLSYEQSRQSMALWTMAAPLMSNDRHISPOAKALLQDKVDVIAINQDPLGKQGR 315

QY 330 QLRQGNFYEYWRPLSGL-----AWAVAMINRQETGGPSRYTIVASLGKGVACNPACFI 384
DB 316 RIYKHKGIEIWSRPITPIYQTYYSYAFVNRRTDGTSPDVAVTRELGL---LISPTGYR 372

QY 385 TQLLPVKRLGFEYWTSLRSHINPTGVLLQL 418
DB 373 VEDLYEEVDYGLSPQTKIKVKNPNSGVVILRAD 406
```

```
RESULT 14
Q4STX0 TETNG
ID Q4STX0 TETNG PRELIMINARY; PRT; 410 AA.
AC Q4STX0;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF14084, whole genome shotgun sequence.
DE (fragment).
GN ORFNames=GSTENG00012743001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
```


RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellini V., Katinka M., Vacherie B.,
RA Blomont C., Skalli Z., Cattolico L., Poullain J., De Betardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.,
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CRAE01014084; CAF95912.1; -; Genomic_DNA.
FT NON TER 410 410
SQ SEQUENCE 410 AA; 46172 MW; F5C661DA72DCB88 CRC64;
Query Match 44.9%; Score 1039; DB 2; Length 410;
Best Local Similarity 48.9%; Pred. No. 3.2e-79;
Matches 207; Conservative 59; Mismatches 116; Indels 42; Gaps 8;
QY 8 LHLGALALRFLALVSWDIPGARALDNLARTPTMGLWHERMCMNLDQCEPDCISEK 67
DB 1 MHLALLLLASVLSL-----GTALLDNLGMLRTTPMGLWLRFCIDCDQDPKNCISEN 54
QY 68 LFMEMALMVSEGWDAGYEYLICDDCWAPQDSRGLQADQRPFGHQLQNAVHVK 127
DB 55 LFTDMADRLSQDGWKLGYVYVNVNIDCWSSKDKKGRLOPDKRPFEGGPKLRYVHNR 114
QY 128 GLKGLIYADVGNKTCAGFGPS-FGYDIDDAQTADNGVLLKFDGCGYCSLENLADGYKH 186
DB 115 GLKGLIYDGMGLTTCGGYGTPLDKIDAKTFAEWEVDMLKFDGCGYSNEVEQ-QQGYPL 173
QY 187 MSALNRTGRSIVYSCWPLMYMPF-QKPNYTIROYCNHWRNFADIDDSWKSILDW 245
DB 174 MSALNATGRPIAYSCWPAYSGLPLPKVNYTQGLCNLWRNYDDIQDSWSDVLNIOW 233
QY 246 TSFNQERIVDVAGFGWGNDDMLVIGNFGLSNQVQTMALWAIMAAPLFMSNDLRHISP 305
DB 234 FEDNQDVIAPAAGRGNDDMLVIGDFGLSQSRTQMALWAIMAAPLFMSNDLRHISA 293
QY 306 QAKALLQDKVIAINODPLGKQYQL-RQGDNFEVWERPLSLGLAWAVMINRQETGGPRS 364
DB 294 EARSILQNKIAIGINQDPLGFGQRLVYKESGIEVFWRLSNDTSALVFFSRR----- 346
QY 365 YTIASVSLGKGVACNPFITQLLPVKRLGFEY-----WTSRLRS-----HINPTG 411
DB 347 -----TDMPIRYKTSLSKLSYPTGYTKIVDVFTQKSSLKDSADFVSVNPTG 394
QY 412 TVL 414
DB 395 VVM 397
RESULT 15
Q9VL27 DROME
ID Q9VL27 DROME PRELIMINARY; PRT; 413 AA.
AC Q9VL27;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE CG5731-PA.
GN Names=CG5731; ORFNames=CG5731;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazef R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosain D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert B.C., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Sten T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Weasman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RT Science 287:2185-2195(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Paclet J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.G.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 10:28:11 ; Search time 5498 Seconds
(without alignments)
4352.691 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MOLRRPELHLCALALRFLA.....RLRSHNPFTGVLLQLENTM 421

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO spool_p/US10602219/runat_23122005_151141_17136/app_query.fasta_1.583
-DB=GenEmbl -QFT=fastcap -SURFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602219@cgn 1 1 7415 @runat_23122005_151141_17136 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.in.*
3: gb.env.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pr.*
9: gb.ro.*
10: gb.sts.*
11: gb.sy.*
12: gb.un.*
13: gb.vl.*
14: gb.hcg.*
15: gb.pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	1266	6	AR653696 Sequence
2	2314	100.0	1266	6	AR654964 Sequence
3	2314	100.0	1278	6	AR653694 Sequence

4	2314	100.0	1278	6	AR654962	Sequence
5	2314	100.0	1284	6	AR653697	Sequence
6	2314	100.0	1284	6	AR654965	Sequence
7	2314	100.0	1290	6	CS135670	Sequence
8	2314	100.0	1290	6	AR653692	Sequence
9	2314	100.0	1290	6	AR654960	Sequence
10	2314	100.0	1290	8	BT006864	Homo sapi
11	2314	100.0	1290	11	AY889631	Synthetic
12	2314	100.0	1290	11	AY889632	Synthetic
13	2314	100.0	1290	11	AY892090	Synthetic
14	2314	100.0	1290	11	BT007835	Synthetic
15	2314	100.0	1296	6	AR653695	Sequence
16	2314	100.0	1296	6	AR654963	Sequence
17	2314	100.0	1306	6	BD182986	Pharmaceu
18	2314	100.0	1306	6	BD185988	Pharmaceu
19	2314	100.0	1308	6	AR653693	Sequence
20	2314	100.0	1308	6	AR654961	Sequence
21	2314	100.0	1312	8	BC002689	Homo sapi
22	2314	100.0	1319	6	AX840234	Sequence
23	2314	100.0	1319	8	HSAGALAR	Human mRNA
24	2314	100.0	1343	6	AR101984	Sequence
25	2314	100.0	1343	6	BD064229	Therapy f
26	2314	100.0	1343	6	BD247365	Treatment
27	2314	100.0	1343	6	CS105027	Sequence
28	2314	100.0	1343	6	AR234439	Sequence
29	2314	100.0	1343	6	AR322462	Sequence
30	2314	100.0	1343	6	AR372138	Sequence
31	2314	100.0	1350	6	CQ899396	Sequence
32	2314	100.0	1350	6	CS033394	Sequence
33	2314	100.0	1350	6	CS035758	Sequence
34	2314	100.0	1350	6	CS042346	Sequence
35	2314	100.0	1350	6	CS044710	Sequence
36	2314	100.0	1352	6	AX523899	Sequence
37	2314	100.0	1390	6	AX780273	Sequence
38	2314	100.0	1393	6	I30313	Sequence 6
39	2314	100.0	1393	6	AX030859	Sequence
40	2314	100.0	1447	6	BD134457	Human nuc
41	2314	100.0	1447	6	AX017286	Sequence
42	2314	100.0	1447	6	AX524990	Sequence
43	2314	100.0	3809	6	AX282869	Sequence
44	2313	100.0	1341	8	AX222627	Homo sapi
45	2310	99.8	1350	6	CQ882046	Sequence

ALIGNMENTS

RESULT 1
LOCUS AR653696 1266 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 11 from patent US 6887696.
ACCESSION AR653696
VERSION AR653696.1 GI:67584281
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Gargier,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 11 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
Location/Qualifiers
1..1266
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:			
Pred. No.:	1.34e-207	Length:	1266
Score:	2314.00	Matches:	421
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

DB: 6 Gaps: 0
US-10-602-219-12 (1-421) x AR653696 (1-1266)

Qy 1 MetGlnLeuArgSerProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGGTTCTCGGACATCTCTGGGCTAGACACTGGACATGGATTGGCAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGAGAGCCA 180
Qy 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGCATCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuHisLeuAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyLeuArgGlnLeu 120
Db 301 GATTCAAGAGGAGAGCTTCAGGACAGACCTCAGCGCTTCTCATGGGATTCGCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAAATTATGTCACAGCAAGGACTGAAGCTAGGAGATTATGTCAGATGTTGGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCCAGCGCTTCCCTGGAGGTTTTGGATACAGCAATGATGCCACAGCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGACTCTGCTAAATTTGATGGTTGTTACTGTGACAGATTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATTAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAAGCAATTTG 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTAACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTGACATTTGATGATTCCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAAATCAGCA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACATGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGCGTAATTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheLeuValTrp 340
Db 961 CAGGACCCCTTGGGCAAGAGGGTACCAGCTTAGCAGGAGACAACCTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360

Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGGACAGGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGGTGGCTGTAACTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263

RESULT 2
AR654964 1266 bp DNA linear PAT 13-JUN-2005
LOCUS AR654964
DEFINITION Sequence 11 from patent US 6890748.
ACCESSION AR654964
VERSION AR654964.1 GI:67586503
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Gargier,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 11 10-MAY-2005;
FEATURES
source location/Qualifiers
1..1266
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1,34e-207 Length: 1266
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6

US-10-602-219-12 (1-421) x AR654964 (1-1266)
Qy 1 MetGlnLeuArgSerProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGGTTCTCGGACATCTCTGGGCTAGACACTGGACATGGATTGGCAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGAGAGAGCCA 180
Qy 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGCATCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuHisLeuAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyLeuArgGlnLeu 120
Db 301 GATTCAAGAGGAGAGCTTCAGGACAGACCTCAGCGCTTCTCATGGGATTCGCAGCTA 360

QY 121 AlaAspTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db 361 GCTAATTATGTTACAGCAAGAGCTGAAGCTAGGATTTATGACAGATGTTGGAAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGCGAGGCTCCCTGGGAGTTTGGATACACGACATTGATGCCACACCTTTGCT 480
 QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
 QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGCTTATAAGCACATGCTCTGTAATAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
 QY 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 601 TCCTGTGAGTGGCTCTTTATATGTTGGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
 Db 661 CAGTACTGCATCATCTGGCGAATTTTGTGACATTTGATGATCTCTGGAAAGTATAAG 720
 QY 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACATGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
 QY 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGln 280
 Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTTCTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGTAAATGGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
 Db 961 CAGGACCCCTTGGGCAAGCAGGCTACCCAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATTAACCCGCGAGGAGTTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
 Db 1141 GCCTGCTTTCATCACACAGCTCTCCCTGTGAAAGGAAAGCTAGGCTTCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCAAGTTAAGAAGTACATAAATCCACAGGCACTGTTTTCTTCAGCTAGTAACAACA 1260
 QY 421 Met 421
 Db 1261 ATG 1263

RESULT 3

AR653694

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

AR653694 Sequence 7 from patent US 6887696.

AR653694

AR653694.1 GI:67584279

Unknown.

ORGANISM

Unknown.

Unclassified.

1 (bases 1 to 1278)

Garger, S.J., Turpen, T.H. and Kumagai, M.H.

Production of lysosomal enzymes in plants by transient expression

Patent: US 6887696-A 7 03-MAY-2005;

Large Scale Biology Corporation; Vacaville, CA

Location/Qualifiers

1..1278

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 1,35e-207 Length: 1278

Score: 2314.00 Matches: 421

Percent Similarity: 100.00% Conservativeness: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0

US-10-602-219-12 (1-421) x AR653694 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20

Db 1 ATGACGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTCTTCTTGGCC 60

QY 21 LeuValSerTyrAspIleProGlyValAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40

Db 61 CTGCTTCTGGACATCCCTGGGCTAGAGCCTGGACACTGGACACTGGATTGGCAAGGACGCT 120

QY 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60

Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATCTGCAACCTTGACTGCGAGGAGGCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80

Db 181 GATTCCTGCATCAGTAGAAGCTCTTCATGAGATGGCAGAGCTCATGCTCTCAGAGGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg 100

Db 241 TGGAAAGGATGAGGTTATGAGTACCTCTGCATTCATGATGATGCTGGTGGCTCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

Db 301 GATTGAGAGGAGCAGACTTCAGGACACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140

Db 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGAGTTATGACAGATGTTGGAAATAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160

Db 421 ACCTGCGAGGCTCCCTGGGAGTTTGGATACACGACATTGATGATGCCACACCTTTGCT 480

QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200

Db 541 GCAGATGCTTATAAGCACATGCTCTTGGCTTGAATAGGACTGGCAGAGGCAATTTGTATC 600

QY 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220

Db 601 TCCTGTGAGTGGCTCTTTATATGTTGGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240

Db 661 CAGTACTGCATCATCTGGCGAATTTTGTGACATTTGATGATCTCTGGAAAGTATAAG 720

QY 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

```

Db      721 AGTATCTTGGACGTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACACGGG 780
Qy      |||
Qy      261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db      |||
Db      781 GGTGGGAATGACCCAGATATGTTAGTGTGGCAACTTTGGGCTCAGCTGGGAATCAGCAA 840
Qy      |||
Qy      281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      |||
Db      841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTTATGCTCTAATGACCTC 900
Qy      |||
Qy      301 ArgHisIleSerProGlnAlaIleValAlaLeuGlnAspLysValIleAlaIleAsn 320
Db      |||
Db      901 CGACACATCAGCCCTCAGCCAAAGCTCTCCTTCAGGATAAGACGTAATTCCTCAAT 960
Qy      |||
Qy      321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAsnAspPheGluValTrp 340
Db      |||
Db      961 CAGGACCCCTTGGGCAAGCAAGGTACCGACTTGTAGACAGGGAGACAACCTTTGAAGTGG 1020
Qy      |||
Qy      341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      |||
Db      1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATTAACCGGACGAGATTTGGT 1080
Qy      |||
Qy      361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      |||
Db      1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
Qy      |||
Qy      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrgluTrpThr 400
Db      |||
Db      1141 GCCTGCTTCATCACACAGCTCTCCTCTGTGAAAGAAAGCTAGGCTTCTATGAATGGACT 1200
Qy      |||
Qy      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db      |||
Db      1201 TCAGGTTAAGAAAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
Qy      |||
Qy      421 Met 421
Db      |||
Db      1261 ATG 1263

RESULT 4
AR654962
LOCUS      AR654962
DEFINITION Sequence 7 from patent US 6890748.
ACCESSION AR654962
VERSION    AR654962.1 GI:67586501
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 1278)
AUTHORS   Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE      Production of lysosomal enzymes in plants by transient expression
JOURNAL    Patent: US 6890748-A 7 10-MAY-2005;
            Large Scale Biology Corporation; Vacaville, CA
FEATURES
            Location/Qualifiers
            1..1278
               /organism="unknown"
               /mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.:      1.35e-207      Length:      1278
Score:          2314.00      Matches:    421
Percent Similarity: 100.00%      Conservatives: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:    0
DB:             6      Gaps:      0

US-10-602-219-12 (1-421) x AR654962 (1-1278)

Qy      1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaIleuArgPheLeuAla 20
Db      |||
Db      1 ATGCAGCTCAGGAACCCAGAACTACATCTGGGCTCGCGCTTGGCTTCTCTGGGCC 60

```

```

Qy      21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db      |||
Db      61 CTCGTTTCTGGGACATCCCTGGGGCTAGACACCTGGACAATGGATTGGCAAGGAGCGCT 120
Qy      |||
Qy      41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db      |||
Db      121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy      |||
Qy      61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db      |||
Db      181 GATTCTCTCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy      |||
Qy      81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db      |||
Db      241 TGGAAAGGATGAGGTTATGATGACTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Qy      |||
Qy      101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db      |||
Db      301 GATTCAGAAGGCAGACTTCAGGACGCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy      |||
Qy      121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db      |||
Db      361 GCTAATATTGTTCAAGCAAGGAGCTGAAGCTAGGGAATTTATGCAAGATGTTGGAATAAAA 420
Qy      |||
Qy      141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db      |||
Db      421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATATACGACATTTGATGCCACGACCTTTGCT 480
Qy      |||
Qy      161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db      |||
Db      481 GACTGGGAGTAGATCTCTAAAATTTGATGGTTGTTACTGTGACAGATTTGGAAAATTTG 540
Qy      |||
Qy      181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db      |||
Db      541 CGAGATGTTTATAAGCACATGTCTCTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGAC 600
Qy      |||
Qy      201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db      |||
Db      601 TCCGTGAGTGGGCTCTTTATATGTGGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy      |||
Qy      221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db      |||
Db      661 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTTGATGATTCCTCGAAAAGATATAAG 720
Qy      |||
Qy      241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      |||
Db      721 AGTATCTTGGACCTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACACGGG 780
Qy      |||
Qy      261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db      |||
Db      781 GGTGGGAATGACCCAGATATGTTAGTGTATGGCAACTTTGGGCTCAGCTGGGAATCAGCAA 840
Qy      |||
Qy      281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      |||
Db      841 GTAACCTCAGATGGGCTCTGGGCTATCATGGCTGCTCCTTTATTTATGCTCTAATGACCTC 900
Qy      |||
Qy      301 ArgHisIleSerProGlnAlaIleValAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db      |||
Db      901 CGACACATCAGCCCTCAGCCAAAGCTCTCCTTCAGGATAAGACGTAATTCCTCAAT 960
Qy      |||
Qy      321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      |||
Db      961 CAGGACCCCTTGGGCAAGCAAGGTACCGACTTGTAGACAGGGAGACAACCTTTGAAGTGG 1020
Qy      |||
Qy      341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      |||
Db      1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATTAACCGGACGAGATTTGGT 1080
Qy      |||
Qy      361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      |||
Db      1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTTCTCTGGGCC 1140
Qy      |||
Qy      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrgluTrpThr 400

```

```
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGCTGAAAAGGAAGCTAGGCTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAGGTTTAAAGTACATATAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263

RESULT 5
AR653697
LOCUS AR653697 1284 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 13 from patent US 6887696.
ACCESSION AR653697
VERSION AR653697.1 GI:67584282
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1284)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 13 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
1..1284
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1,36e-207 Length: 1284
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-219-12 (1-421) x AR653697 (1-1284)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTTCTCATGGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGATGCTTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGCTGAGCTAGGATTTATGAGATTGGAATATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTGATGTCAGACCTTTGCT 480
```

```
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTTACTGTGACAGTTTGAAAAATTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGCAATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCTCTTTATATGCTGGCCCTTCAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGCGAAATTTTCTGCATTTGATGATCTCTGCGAAAGTATAAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAATTTGTGATGTTGCTGGACCAAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGCCAGATATGTTAGTGAATTTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTATTCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGCTTAATTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCGAAGCAAGGGTACCAAGCTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTAGCTATGATAAACCAGCAGAGATTGCT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAGAAGAGCTAGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAGGTTTAAAGTACATATAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263

RESULT 6
AR654965
LOCUS AR654965 1284 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 13 from patent US 6890748.
ACCESSION AR654965
VERSION AR654965.1 GI:67586504
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1284)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 13 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
Location/Qualifiers
```

```
source 1..1284
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1.36e-207 Length: 1284
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-219-12 (1-421) x AR654965 (1-1284)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTCGCTTCCTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGGATTGGCAGGAGCGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTCAACCTTGACTGCCAGGAAGAGCA 180

QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTyrGlnTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGAGGTTATGAGTACTCTGCATGTATGACTGTGGATGGCTCCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTTCAGAAGGCAGACTTCAGGCAGACCTCAGCGCTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATTATGTTTCACAGCAAGAGCTCAGAGCTAGAGCTAGGATTTATGCAGATGTTTGGAAATAAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACTTCGGCAGGCTTCCTGGGAGTTTGGATNACTAGCATTCAGATTCGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGATTTGGAAAAATTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 CGAGATGGTTATPAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCTCTGAGTGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAACAGGAGAGAAATTTGTGTGTTGTCTGACAGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GTTTGGATGACCCAGATATGTATGTATGTGGCACTTTGGCTCAGCTGGGATCAGCAA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGTCTCTTTTATTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGAGTAAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTTGGGCAAGCAAGGTTACCCAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCCCTGAGCTATGATATAAACCCGAGAGAGATTGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValaLysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCTTGTAAATCT 1140
QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
QY 421 Met 421
DB 1261 ATG 1263

RESULT 7
CS135670 1290 bp DNA linear PAT 09-AUG-2005
LOCUS Sequence 1 from Patent WO2005069192.
DEFINITION CS135670
ACCESSION CS135670
VERSION CS135670.1 GI:72056330
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE
1 Garman,S.C.
Crystal structure of human a-galactosidase
Patent: WO 2005069192-A 1 28-JUL-2005;
Transkaryotic Therapies, Inc. (US); Selden, Richard F. (US);
Garboczi, David N. (US); Treco, Douglas A. (US); NIH/NIHID (US);
Garman, Scott C. (US); Borowski, Marianne (US); Kinoshita, Carol M.
(US)
FEATURES
Location/Qualifiers
source 1..1290
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 1.37e-207 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-219-12 (1-421) x CS135670 (1-1290)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTCGCTTCCTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTyrGlnTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGAGGTTATGAGTACTCTGCATGTATGACTGTGGATGGCTCCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTTCAGAAGGCAGACTTCAGGCAGACCTCAGCGCTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATTATGTTTCACAGCAAGAGCTCAGAGCTAGAGCTAGGATTTATGCAGATGTTTGGAAATAAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACTTCGGCAGGCTTCCTGGGAGTTTGGATNACTAGCATTCAGATTCGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGATTTGGAAAAATTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 CGAGATGGTTATPAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCTCTGAGTGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAACAGGAGAGAAATTTGTGTGTTGTCTGACAGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GTTTGGATGACCCAGATATGTATGTATGTGGCACTTTGGCTCAGCTGGGATCAGCAA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGTCTCTTTTATTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGAGTAAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTTGGGCAAGCAAGGTTACCCAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCCCTGAGCTATGATATAAACCCGAGAGAGATTGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValaLysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCTTGTAAATCT 1140
QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
QY 421 Met 421
DB 1261 ATG 1263
```


Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCAGCTGGACAATGGATTGGCAAGGAGCGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTCATGCTGCCAGGAGGCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGlnTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGGAAGGATGAGGTATAGTACTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAGGCAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGTGGATTGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGATTTATGCAAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGGAGGCTTCCCTGGGAGTTTGGATACAGACATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGTAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGCCCTCTTTATATGTGGCCCTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGACATCTTTTAACCAAGGAGAGAAATTTGTGATTTGTCTGGACACGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTTATTCATGCTAAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGCTTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAGGGTACCAGCTTTAGACAGGGAGACAATTTGAAAGTGGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCCGACAGGAGTTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGACTGTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTTCATCACACAGCTCTCCCTGTGAAAGGAGGAGCTAGGGTTCTATGAATGGACT 1200

Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAGTTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTAGAAAATACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
RESULT 8
LOCUS AR653692 1290 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 3 from patent US 6887696.
ACCESSION AR653692
VERSION AR653692.1 GI:67584277
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 3 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
1. .1290
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1.37e-207 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-602-219-12 (1-421) x AR653692 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCAGCTGGACAATGGATTGGCAAGGAGCGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGTACTGCCAGGAGGCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGlnTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGGAAGGATGAGGTATAGTACTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAGGCAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGTGGATTGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGATTTATGCAAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGGAGGCTTCCCTGGGAGTTTGGATACAGACATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

Db	481	GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTGTGACAGATTTGGAAATTTG	540
Qy	181	AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr	200
Db	541	GCAGATGGTTATTAAGCACATGTCCTTGGCCCTGAATAGACTGGCAGAACATTTGTATC	600
Qy	201	SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg	220
Db	601	TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA	660
Qy	221	GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
Db	661	CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG	720
Qy	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
Db	721	AGTATCTTGGACTGGACATCTTTTAACCCAGGAGAGAAATTTGTGATTTGCTGGACAGGG	780
Qy	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln	280
Db	781	GGTTGGAATGACCCAGATATGTAGTGTGGCAACTTTTGGCCTCAGCTGGAAATCAGCAA	840
Qy	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
Db	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCTAATGACCTC	900
Qy	301	ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn	320
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATTAAGGACGTAATTTGCCATCAAT	960
Qy	321	GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
Db	961	CAGGACCCCTTGGCAAGCAAGGATACCGCTTACAGCGAGAGACAACCTTTGAAGTGTGG	1020
Qy	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
Db	1021	GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAACCGCGCAGGATTTGT	1080
Qy	361	GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
Db	1081	GGACCTCGCTCTATACCATCGAGTTGCTTCCCTGGGTAAAGAGATGGCCCTGTAATCT	1140
Qy	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr	400
Db	1141	GCCTGCTTCATCACAGCTCTCCTCTGTAAGAGAGACTAGGGTTCATGATGGACT	1200
Qy	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr	420
Db	1201	TCAGGTTAAGAGTCACATAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA	1260
Qy	421	Met 421	
Db	1261	ATG 1263	
RESULT 9			
AR654960			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			

ORIGIN		/mol_type="genomic DNA"	
Alignment Scores:			
Pred. No.:	1.37e-207	Length:	1290
Score:	2314.00	Matches:	421
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-10-602-219-12 (1-421) x AR654960 (1-1290)			
Qy	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC	60
Qy	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
Db	61	CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT	120
Qy	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	121	ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	180
Qy	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	181	GATTCCTGCATCAGTAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC	240
Qy	81	TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg	100
Db	241	TGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAGA	300
Qy	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
Db	301	GATTCAGAAAGCAGACTTCAGCGACACCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA	360
Qy	121	AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys	140
Db	361	GCTAATATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTTGGAAATAAA	420
Qy	141	ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla	160
Db	421	ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATTCAGCACTGGCCAGACCTTTGCT	480
Qy	161	AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu	180
Db	481	GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTTG	540
Qy	181	AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr	200
Db	541	GCAGATGGTTATAGCACATGTCCTTGGCCCTGAATAGACTGGCAGAGCAATTTGTGTAC	600
Qy	201	SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg	220
Db	601	TCCTGTGAGTGGCCCTCTTATATGTGGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA	660
Qy	221	GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
Db	661	CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTGGAAAGATATAAG	720
Qy	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
Db	721	AGTATCTTGGACTGGACATCTTTTAACCCAGGAGAGAAATTTGTTGATGTTGTCGACCGG	780
Qy	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln	280
Db	781	GGTTGGAATGACCCAGATATGTTAGTGTGGCAACTTTGGCCTCAGCTGGAAATCAGCAA	840
Qy	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
Db	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCTAATGACCTC	900


```

Qy 301 ArgHisIleSerProGlnAlaIysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGAGCATTAATGGCCATCAAT 960

Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGGACAGCAGGATACAGCTTAGACAGGGAGACACTTTGAAGTGTGG 1020

Qy 341 GlnArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGlnIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCAGCAGGAGATTGGT 1080

Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTCTGCTTCCCTGGGTAAGGAGTGGCCCTGTAATCCT 1140

Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGlnTyrThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCTGTGAAAAGAGAGCTAGGGTTTCTATGAATGGACT 1200

Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTACATATAATCCACAGGCACTGTTTGTCTTCAGTAGAAAACACA 1260

Qy 421 Met 421
Db 1261 ATG 1263

RESULT 10
LOCUS BT006864 1290 bp mRNA linear PRI 13-MAY-2003
DEFINITION Homo sapiens galactosidase, alpha mRNA, complete cds.
ACCESSION BT006864
VERSION BT006864.1 GI:30582566
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1290)
Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
Cloning of human full-length cDNAs in BD Creator(TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1290)
Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This cDNA clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide Kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
Location/Qualifiers
1..1290
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GH00112X1.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"

```

```

/note="Vector: pDNR-Dual"
1..1290
/codon_start=1
/product="galactosidase, alpha"
/protein_id="AAP35510.1"
/db_xref="GI:30582567"
/translation="MQLRNPRLHGLGALALRFLALVSWDIPGARALDNLGLARTPTMGW
LHFRFNCNLCQEPSCISERKLVHSEGMWADAGTYELICIDCWNPQRDS
EGRLQADPQRPFGIRQLANYVHSGKLGIYADGNKTCAGFPGSFYDIDAQTF
DWGVDLKFDCGYCDSLENLADYGRHSLNRTGRSIVYSCENFLYVMPFQKPYTE
IROYCNHRNFADIDDSKSIILDMTSFNQERIVDVAGPGWNDPMLVTIGNFGLS
WNOQVOTOMALWAINAAPLFMSNDRIHLSPOAKALLQDKDLAINQDPLGKQGYQLRQG
DNFVWRPDLGLAWAVAMINRQISGRPSYTIIVASLKGVCACNPACFITQLLPVKR
KLGYEWTSLRSLRSHINPTGVLLQLENTQMQLSKDLL"
ORIGIN
Alignment Scores: 1.37e-207 Length: 1290
Pred. No.: 2314.00 Matches: 421
Score: 2314.00 Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 8
US-10-602-219-12 (1-421) x BT006864 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGCGCGCTTGGCGCTTCTCGTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGACACTGGCAATGGATTGGCAAGAGCGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Qy 61 AspSerCysIleSerGlyLeuPheMetGluMetAlaGluLeuMetValSerGlyGly 80
Db 181 GATTCCTGCATCAGTGAAGCTCTTTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGlyTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGAGGTTATGAGTACCTCTGCACTTGTGATGATGCTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGlyValArgLeuGlnAlaAspProGlnArgPheProHisGlyLeuArgGlnLeu 120
Db 301 GATTGAGGAGGAGACATTCAGGAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTCAGCAAGAGGACTGAAGCTAGGATTTATGCAGATGTTTGGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGTACTACGACATTTGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGGATAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAGCACATGTCCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTCGAAAGATTAAG 720

```

QY 241 SerIleLeuAspTrpThrSerPhehngInGluAtrgIleValaspValaIaGlyProGly 260
 Db |||||
 721 AGTATCTGGACTGGACATCTTTTAAACGAGGAGAAATTGTGTGCTGGACAGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPhGlyLeuSerTrpAsnGln 280
 Db |||||
 781 GGTGGATGACCCAGATATGTAGTGATGGCACTTTGGCTCAGCTGGATCAGCA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaIaProLeuPheMetSerAsnAspLeu 300
 Db |||||
 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTATGTCTAATGACCTC 900

QY 301 ArgHisIleSerProGlnAlaIysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db |||||
 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTGCCATCAAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuAtrgGlnGlyAsnPhGluValTrp 340
 Db |||||
 961 CAGGACCCCTTGGGCAAGCAGGGTACCAGCTTAGACAGGAGACAACTTTCAGATGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnAtrgGlnGluIleGly 360
 Db |||||
 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATCATATAACCGGCAGGAGATTGCT 1080

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db |||||
 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAAGGAGTGGCCCTGTATCTCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysAtrgLysLeuGlyPheTyrGluTrpThr 400
 Db |||||
 1141 GCTGCTTCATCACAGCTCTCTCTGTGAAAGGAGCTAGGGTTCTATGANTGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnGlnLeuGluAsnThr 420
 Db |||||
 1201 TCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAATACA 1260

QY 421 Met 421
 |||
 Db 1261 ATG 1263

RESULT 11
 AY889631

LOCUS
 DEFINITION Synthetic construct Homo sapiens clone FLH025769.01X galactosidase alpha (GLA) mRNA, complete cds.

ACCESSION
 VERSION AY889631.1 GI:60655716
 KEYWORDS Human ORF Project.
 SOURCE synthetic construct
 ORGANISM other sequences; artificial sequences.

REFERENCE
 1 (bases 1 to 1290)
 Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F., Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E., Williamson, J., and Labaer, J.
 Cloning of human full-length CDS in Creator (TM) recombinational vector system
 Unpublished

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

COMMENT
 This CDS clone is a part of a collection of human full-length expression clones generated by Harvard Institute of Proteomics. This ORF clone has been cloned with normalized stop-codon. The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pONR-Dual vector. Additional sequences in the clone: 'ACC' after SalI site and

before 'ATG' to provide Kozak consensus sequence. Each clone is clonally isolated and full-length sequence-verified.

FEATURES

source

1. .1290

/organism="synthetic construct"

/mol_type="mRNA"

/db_xref="taxon:32630"

/clone="FLH025769.01X"

/lab_host="Escherichia coli DH5alpha Tl resistant"

/note="derived from MGC template"

1. .1290

/gene="GLA"

1. .1290

/gene="GLA"

/codon_start=1

/transl_table=11

/product="galactosidase alpha"

/protein_id="AA32422.1"

/db_xref="GI:60655717"

/translation="MQLNPHELHLCALALRFLALVSWDIPGARALDNGLARTPTMGW
 LHWRFMCLDQEPDSCISEKLPMENAEMLVSEGWKADGVEYLICIDDCWAPQDSDS
 EGRLOADQRFPHGIRQLANYVHSGKLKGIYADVGNKTCAGFGSGYIDIDAQTFP
 DWGVDLLKFDGCDLSLENLADGYKMSLALNRGRSIVYSCENLYMWPFOKPNYTE
 IROYCNWRNFADIDDSWSKISILDMTISFNQERIVDVAGPGWMDPMLVIGNFGLS
 WNOQYVOMALWMAAPLFMSNDLHISFOAKALQDKDVIATINQDPLGKQGYQLRQG
 DNFEWFERPLSLAWAVAMINROEIGGPRSYTIAVASLGKGVACNPACFITQLLPVKR
 KLGFYEWTSRLRSHINPTGTVLLQLLENTMQMSLKDLL"

ORIGIN

Alignment Scores:

Pred. No.: 1.37e-207 Length: 1290
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 11 Gaps: 0

US-10-602-219-12 (1-421) x AY889631 (1-1290)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db |||||
 1 ATGCAGCTCAGGAACCCAGAACTACATCTGGGCTGCGGCTTTCGGCTTCCTCTGGGC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db |||||
 61 CTCGTTCCTCGGACATCCTCGGGCTAGACACCTGGACAAATGGATTCGCAAGGACGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db |||||
 121 ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTCAACCTTGACTCCAGGAAGGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db |||||
 181 GATTCTCTGCATCAGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnAtrg 100
 Db |||||
 241 TGGAAGGATGCAAGGTATGAGTACCTCTGCATTGATGACTGTGTGATGGCTCCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnAtrgPheProHisGlyIleArgGlnLeu 120
 Db |||||
 301 GATTCCAGAGGCAGACTTCAGGCAGACCCCTCAGCGGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db |||||
 361 GCTAATATGTTTTCACAGCAAGGAGCTGAGAGTATGAGGATTTATGCAGATGTGGAATAAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 Db |||||
 421 ACTTCGCGAGGCTTCCTCGGAGTTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db |||||

```
Db 481 GACTGGGAGTAGATCTCTCTAAATTTGATGGTTGTTACTGTGACAGCTTTGCAAAATTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
Db 541 CGAGATGTTATAGACATGTCCTTGGCCCTGAATAGACTGGCAGAAAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCATCACTGGCGAAATTTTCTGACATGTATGATTTCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTATGTTGCTGGACAGGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGln 280
Db 781 GGTGGAAATGACCCAGATATGTAGTATGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCCAAGCTCTCTTCAGGATAGGAGTAAATGGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGGCAAGAGGTATCCAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnIleGly 360
Db 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATGATAAACCAGCAGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAAGTAGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCNAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAATACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
RESULT 12
LOCUS AY889632
DEFINITION Synthetic construct Homo sapiens clone FLH025770.01x galactosidase
alpha (GLA) mRNA, complete cds.
ACCESSION AY889632
VERSION AY889632.1 GI:60655718
KEYWORDS Human ORF project.
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
Williamson,J. and LaBaer,J.
Cloning of human full-length CDS in Creator (TM) recombinational
vector system
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1290)
```

```
AUTHORS Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-2005) Biological Chemistry and Molecular
Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
Cambridge, MA 02141, USA
COMMENT This CDS clone is a part of a collection of human full-length
expression clones generated by Harvard Institute of Proteomics.
This ORF clone has been cloned with normalized stop-codon. The CDS
has been directionally cloned using BD In-Fusion(TM) cloning system
between the Sali and HindIII sites of the pDNR-Dual vector.
Additional sequences in the clone: 'ACC' after Sali site and
before 'ATG' to provide Kozak consensus sequence. Each clone is
clonally isolated and full-length sequence-verified.
FEATURES
Location/Qualifiers
1..1290
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="FLH025770.01x"
/lab_host="Escherichia coli DH5alpha T1 resistant"
/note="derived from MGC template"
gene 1..1290
/gene="GLA"
CDS 1..1290
/gene="GLA"
/codon_start=1
/transl_table=11
/product="galactosidase alpha"
/protein_id="AA32423.1"
/db_xref="GI:60655719"
/translacion="MQLRNPGLHLCALALFLALVSWDIPGARALDNLARTPTMGW
LHWRPNCNLDQCEPDCISEKLFMEAMELMWSGWDAGVEYLICDDCWAPQDTS
EGRIQADPQPPHGIROLANVYHSLKGLGIYADVGNKTCAGFPGSFGYDIDAOTFA
DWGYDLKFGCYCDSLENLADGYKMSLANRTGRSLVYSCWPLNMPKPNVTE
IRQCNWRFPADIDDSKWSIKSLDWTSEFQERIVDVAGGGWNPDMVLVGNFGLS
WNQQTQWALWAIWAALFLMNSDLRHSPOKALLQDKDVIAINDPDKGQGYLQKQG
DNFVWRPLSGLAWAMINIRBGIPRSYTIASVSLGKGVACNPACFITOLLVPRK
KLGFYEWTSRLRSHINPTGTVLLQLENTMQMSLKDLL"
ORIGIN
Alignment Scores:
Pred. No.: 1.37e-207 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11
US-10-602-219-12 (1-421) x AY889632 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCGCTGGCGCTTTCCTCTCTCTGCCC 60
Qy 21 LeuValSerTyrAspIleProGlyValAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGACATCTCCCTGGGCTAGACACTGGACATGGATTGGCAAGGAGCCT 120
Qy 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCGCAGAGAGCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGTCTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg 100
Db 241 TGGAAAGGATGAGGTTATGATGACTCTGCTTGTGATGATGCTTGGTGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
```

Db 301 GATTTCAGAGGAGCAGCTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAenTyrValHisSerIysGlyLeuIysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTCAGCAGCAAGAGCTGAAGCTAGGAGATTTATGTCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTCGGAGTTTTGGATACTACGACATGTGATGCCAGACCTTTGCT 480
Qy 161 AspTyrGlyValAspLeuLeuIysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGCTGTCAGTGCAGATTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATPAGCACATGCTTCGGCCCTGAATAGACTGGCAGAAAGCAATGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGCCCTCTTTATATGTGGCCCTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrParghenPheAlaAspIleAspAspSerTyrIysSerIleLys 240
Db 661 CAGTACTGCATCACTGGCGAAATTTTGTGCATATTGATGATTCCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACCAAGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTTGTCACACTTTGGCTCAGCTGGATCAGCA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaIaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGCCCTCTGGGCTATCATGCTGCTCTCTTTATTTCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 GCACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATPAGGACGTAAATGGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGCAGCAGGATACCGCTTAGACGGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGCAGGAGATTGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCCAGTTGCTTCCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCTGTGAAAGGAGCTAGGGTTCTATGAAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTTCCTTCAGCTAGAAAATACA 1260
Qy 421 Met 421
Db 1261 ATG 1263

RESULT 13
AY892090
LOCUS AY892090 1290 bp mRNA linear SYN 29-MAR-2005
DEFINITION Synthetic construct Homo sapiens clone FLH025766.01L galactosidase
alpha (GLA) mRNA, partial cds.
ACCESSION AY892090

AY892090.1 GI:60652624
Human ORF project.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1290)
AUTHORS Hines,B., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
TITLE Cloning of human full-length CDS in Creator (TM) recombinational
vector system
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1290)
AUTHORS Hines,B., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-2005) Biological Chemistry and Molecular
Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
Cambridge, MA 02141, USA
COMMENT This CDS clone is a part of a collection of human full-length
expression clones generated by Harvard Institute of Proteomics.
This ORF clone has been cloned without stop-codon (to allow fusion
with C-terminal tag). The CDS has been directionally cloned using
BD In-Fusion(TM) cloning system between the SalI and HindIII sites
of the pDNR-Dual vector. Additional sequences in the clone: 'ACC'
after SalI site and before 'ATG' to provide Kozak consensus
sequence; 'GG' after last codon and before HindIII site to maintain
reading frame. Each clone is clonally isolated and full-length
sequence-verified.
FEATURES
source
1. 1290
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="FLH025766.01L"
/lab_host="Escherichia coli DH5alpha T1 resistant"
/note="derived from MGC template"
gene
1. >1290
/gene="GLA"
CDS
1. >1290
/gene="GLA"
/codon_start=1
/transl_table=11
/product="galactosidase alpha"
/protein_id="AA029007.1"
/db_xref="GI:60652625"
/translation="MQLRNPHELHGCALALFLALVSWDIPGARALDNLARTPTMGW
LHWRFMCNLDQEPDSCISEKLFMEAMLMVSEGWDAGYEYLICDDCWAPQDTS
EGRLQADPQRPHGIROLANTYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTA
DWGVDLLKFDGDCYCDLSLNLADGKXMSLALNRTGRSIVYSCFWLWMPFKPNYTE
IRQVCHWRNFADIDDSKSIKSLDWTSTFNOERIVDVAGGGWNPDMVLVGNGLS
WNQVQTQALWAIWAAPLUFMSNDLRHISFQAKLLQDKDVIATNQDPLGKQGYLQRQ
DNFEWPERPLSLAWAMINRQIGGPRSVIAVASLGKGVACNPACFTITQLLPVKR
KLGFVETWSRLSRSHINPTGTVLLOLENTMQMSLKDLLL"
ORIGIN
Alignment Scores:
Pred. No.: 1.37e-207 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Db: 11 Gaps: 0
US-10-602-219-12 (1-421) x AY892090 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGACTACATCTGGGCTCGGCGCTTGGCTTCGCTTCCTGGCC 60
Qy 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40

Db 61 CTCGTTTCTCGGACATCCCTGGGGCTAGAGCAGCTGGACAATGGATTGGCAAGAGCGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGTCATCAGTGAGAAGCTCTTCATGTGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleLeuAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGACAGGTATAGTACTCTGCAATGTGATGCTGTGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleAArgGlnLeu 120
Db 301 GATTGAGAAGCAGCAGCTTCAGGAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaSerTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGCTGAAGCTAGGAGATTTATGCAGATGTTTGGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTTGGATACTACGACATGTATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTG 600
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTCAAAAGCCCAATATATACAGAAATCGA 660
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATCTGCGGAAATTTTCTGCATTTGATGATTTCTGGAAAGTATAAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTAAACAGGAGAGAAATTTGTTGTTGCTGGCCAGCGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTATTATGCTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCCAAGCTCTCCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTAGACAGGGAGAGCAACTTTGAAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTTGGT 1080
Qy 361 GlyProArgSerTrpIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTTCATCACAGCTCTCCCTGTGAAAGGAGAGCTAGGCTTCTATGAATGGACT 1200

Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTAAAGAGTACATAAATCCACAGGCACCTGTTTGTCTTTCAGTAGAAAATACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
RESULT 14
LOCUS BT007835 1290 bp mRNA linear SYN 13-MAY-2003
DEFINITION Synthetic construct Homo sapiens galactosidase, alpha mRNA, partial cds.
ACCESSION BT007835
VERSION BT007835.1 GI:30584508
KEYWORDS FLI CDNA.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1290)
AUTHORS Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A.
Cloning of human full-length CDSs in BD Creator (TM) System Donor vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1290)
AUTHORS Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A.
Direct Submission
TITLE Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame.
Clone distribution: <http://bioinfo.clontech.com/orfclones>.
FEATURES
Location/Qualifiers
1..1290
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH00112L1.0"
/clone_lib="BD Creator (TM) CDS Library derived from MGC collection"
/lab_host="DHSalpha T1 resistant"
/note="vector: pDNR-Dual"
1..>1290
/note="Mutations: 1289:Stop->Leu"
/codon_start=1
/transl_table=1
/product="Homo sapiens galactosidase, alpha"
/protein_id="AAP36507.1"
/db_xref="GI:30584509"
/translation="MQLRNPELHIGCALALRFLALVSWDIFGARALDNLGARTPTMGW LHWRFNCLDCEPDCISEKLFMELMSEGVKDGVEYLICIDCWAPORDS EGRLOADPQPPHGIROLANYVHSGKLGIYADVGNKTCAGFGSGYDIDAQTEA DWGVLLKFGGCTCDSLENADGYKHMSLANRGRSIVSYCEPNLWMPQKPNYTE IROYCNHRNFADIDDSWKSITKSLDTSFNQRIVDVAGPGGNDPMLVIGNFGLS NQQVQTQMLWALMAAPLMNSDRHISPOKALLQDKDVTAINDPDKGQYLRQG DNFVWRPLSGLAWAMINRQISGIPRSYTIIVASLIGKGVACNPACFITQLLPVKR KLGFEVTSRLRSHINPTGTVLQLQENTMQMSLKDLLL"
ORIGIN
Alignment Scores:
Pred. No.: 1.37e-207 Length: 1290

Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0

US-10-602-219-12 (1-421) x BT007835 (1-1296)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTCGCTTCCTGGCC 60

Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCATTGGACATGGATTGGCAAGAGCGCT 120

Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCA 180

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGAGGTTATGAGTACTCTGCATTGATGACTGTGGATGGCTCCCCAAGA 300

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAAGCAGACTTCAGGCAGACCTCAGCGCTTCTCTCATGGGATTCGCCAGCTA 360

Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTCACAGCAAGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAATAA 420

Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACTACGACATGTGATGCCAGACCTTGGT 480

Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTGTACTGTGACAGATTTGGAAAAATTG 540

Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 CGAGATGGTTATGAGCACATGCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600

Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGCCCTTTTATATGTGGCCCTTCAAGGCCCAATATACAGAAATCCGA 660

Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAATATAAG 720

Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnLysArgIleValAspValaGlyProGly 260
Db 721 AGTATCTGGACTGGACATCTTTTAAACCAGGAGAGAAATTTGTGATGTTGCTGGACAGGG 780

Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGATGACCCAGATATGTAGTATGGCAACTTTGGCTCAGCTGGATCAGCNA 840

Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900

Qy 301 ArgHisIleSerProGlnLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960

Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340

Db 961 CAGGACCCCTTTGGGCAAGAGGTTACAGCTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020

Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValaMetIleAsnArgGlnGluIleGly 360

Db 1021 GAAGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCCGCGAGAGATTGGT 1080

Qy 361 GlyProArgSerTyrThrIleAlaValaAlaSerLeuGlyLysGlyValaAlaCysAsnPro 380

Db 1081 GGACTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTATATCT 1140

Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400

Db 1141 GCCTGCTTATCATCACAGCTCTCCCTGTGAAAAGGAAGCTAGGCTTCTATGAATGGACT 1200

Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420

Db 1201 TCAAGGTTAAGAGTTCACATAAATCCCAAGGCACCTGTTTTGCTTTCAGCTAGAAATACA 1260

Qy 421 Met 421

Db 1261 ATG 1263

RESULT 15

AR653695 LOCUS AR653695 1296 bp DNA linear PAT 13-JUN-2005

DEFINITION Sequence 9 from patent US 6887696.

ACCESSION AR653695

VERSION AR653695.1 GI:67584280

KEYWORDS SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1296)

AUTHORS Garg,S.J., Turpen,T.H. and Kumagai,M.H.

TITLE Production of lysosomal enzymes in plants by transient expression

JOURNAL Patent: US 6887696-A 9 03-MAY-2005;

Large Scale Biology Corporation; Vacaville, CA

FEATURES

source 1..1296

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 1,388-207 Length: 1296

Score: 2314.00 Matches: 421

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0

US-10-602-219-12 (1-421) x AR653695 (1-1296)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20

Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTCGCTTCCTGGCC 60

Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40

Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCATTGGACATGGATTGGCAAGAGCGCT 120

Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60

Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 180

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80

Db 181 GATTCCTGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100

Db 241 TGGAGGATGAGGTTATGAGTACTCTGCATTGATGACTGTGGATGGCTCCCCAAGA 300


```
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db |||||
Db 301 GATTGAGAAGGCAGACATTCAGGCAGACCCCTCAGGCGCTTTCCTCATGGGATTGCGCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db |||||
Db 361 GCTAATTATGTTACAGCAAGAGGACTGAAGCTAGGGATTTATGAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db |||||
Db 421 ACCTGGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATTTGATGCCACACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db |||||
Db 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db |||||
Db 541 GCAGATGGTTATAAGCATGTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db |||||
Db 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db |||||
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCATTTGATGATTCCTGGAAAAATATAAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db |||||
Db 721 AGTATCTGGAGCTGGACATCTTTAAACCAGGAGAGATTTGTGATTTGCTGGACCCAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db |||||
Db 781 GGTGGATGACCCAGATATGTTAGTGTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaIaProLeuPheMetSerAsnAspLeu 300
Db |||||
Db 841 GTAACCTCAGATGGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaIysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db |||||
Db 901 CGACACATCACCCCTCTCAAGCCAAAGCTCTCCTTCAGGATAGGACGTAAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db |||||
Db 961 CAGGACCCCTTGGGCAAGCAGGGTACCAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db |||||
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCCGGCAGGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db |||||
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db |||||
Db 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db |||||
Db 1201 TCNAGGTTAAGAAGTCACATAAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAAACACA 1260
Qy 421 Met 421
Db |||
Db 1261 ATG 1263
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 08:58:24 ; Search time 44 Seconds
(without alignments)
791.056 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQLNPELHGCALALRFLA.....RLRSHINPTGTVLLQLENTM 421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5-COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/6-COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/H-COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCITUS-COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/RE-COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	421	2	US-09-993-059-12
2	2314	100.0	421	2	US-10-103-327-12
3	2314	100.0	427	2	US-09-993-059-14
4	2314	100.0	427	2	US-10-103-327-14
5	2314	100.0	429	1	US-07-602-824A-2
6	2314	100.0	429	1	US-07-602-608-3
7	2314	100.0	429	1	US-07-983-451-2
8	2314	100.0	429	1	US-08-261-578-3
9	2314	100.0	429	1	US-08-261-577-7
10	2314	100.0	429	1	US-08-261-577-9
11	2314	100.0	431	2	US-09-993-059-10
12	2314	100.0	431	2	US-10-103-327-10
13	2314	100.0	435	2	US-09-993-059-6
14	2314	100.0	435	2	US-10-103-327-6
15	2310	99.8	429	2	US-09-070-356-4
16	2309	99.8	424	2	US-09-993-059-8
17	2309	99.8	424	2	US-10-103-327-8
18	2309	99.8	428	2	US-09-993-059-4
19	2309	99.8	428	2	US-10-103-327-4
20	2296	99.2	428	6	5179023-4
21	2293	99.1	417	2	US-09-993-059-16
22	2293	99.1	417	2	US-10-103-327-16
23	2293	99.1	423	2	US-09-993-059-18
24	2293	99.1	423	2	US-10-103-327-18
25	2226	96.2	409	2	US-09-993-059-22
26	2226	96.2	409	2	US-10-103-327-22
27	2213	95.6	401	2	US-09-993-059-20

28	2213	95.6	401	2	US-10-103-327-20	Sequence 20, Appl
29	2152	93.0	390	2	US-09-176-666-7	Sequence 7, Appl
30	2152	93.0	391	2	US-09-176-666-6	Sequence 6, Appl
31	2152	93.0	392	2	US-09-176-666-5	Sequence 5, Appl
32	2152	93.0	393	2	US-09-176-666-4	Sequence 4, Appl
33	2152	93.0	394	2	US-09-176-666-3	Sequence 3, Appl
34	2152	93.0	396	2	US-09-176-666-2	Sequence 2, Appl
35	2152	93.0	398	2	US-08-928-881-26	Sequence 26, Appl
36	2152	93.0	398	2	US-09-176-666-1	Sequence 1, Appl
37	2152	93.0	398	2	US-09-543-921-26	Sequence 26, Appl
38	2152	93.0	398	2	US-09-266-014-4	Sequence 4, Appl
39	2152	93.0	398	2	US-09-491-759-26	Sequence 26, Appl
40	2152	93.0	398	2	US-10-360-101-202	Sequence 202, App
41	2147	92.8	389	2	US-09-176-666-8	Sequence 8, Appl
42	2142	92.6	388	2	US-09-176-666-9	Sequence 9, Appl
43	2136	92.3	387	2	US-09-176-666-10	Sequence 10, Appl
44	2131	92.1	385	2	US-09-176-666-11	Sequence 11, Appl
45	2110	91.2	381	2	US-09-176-666-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-12
; Sequence 12, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-12

Query Match	100.0%	Score 2314;	DB 2;	Length 421;
Best Local Similarity	100.0%	Pred. No. 1.4e-234;		
Matches 421;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MQLNPELHGCALALRFLALVSWDIPGARALDNLARTPTMGMLHWRFCNLDQCEP	60	
Db	1	MQLNPELHGCALALRFLALVSWDIPGARALDNLARTPTMGMLHWRFCNLDQCEP	60	
Qy	61	DSCLSEKLFMEASLMYSEGHKADAGEYLCIDDCWAPORSEGLQADPQPFPHGIQRL	120	
Db	61	DSCLSEKLFMEASLMYSEGHKADAGEYLCIDDCWAPORSEGLQADPQPFPHGIQRL	120	
Qy	121	ANYVHSKGLGIYADVGNKTCAGPFGSGFYVDIDAOQFADGWVDLLKFDGCGYCDSENL	180	
Db	121	ANYVHSKGLGIYADVGNKTCAGPFGSGFYVDIDAOQFADGWVDLLKFDGCGYCDSENL	180	
Qy	181	ADGKHSMLALNRTGRSIVYSCWPLMYMPFKENYTEIRQYCNHWRNFADIDDSWKSIIK	240	
Db	181	ADGKHSMLALNRTGRSIVYSCWPLMYMPFKENYTEIRQYCNHWRNFADIDDSWKSIIK	240	
Qy	241	SILDWTSFNQRIYDVVAGPGGWNPDMLVINGFGLSNQVQVTOMALWAIMAAPLPMNDL	300	
Db	241	SILDWTSFNQRIYDVVAGPGGWNPDMLVINGFGLSNQVQVTOMALWAIMAAPLPMNDL	300	
Qy	301	RHISPOKALLQDKDVIAINODPLGKQYQLRGQDNFEVWERPLSGLAWAVAMINRQIIG	360	
Db	301	RHISPOKALLQDKDVIAINODPLGKQYQLRGQDNFEVWERPLSGLAWAVAMINRQIIG	360	
Qy	361	GPRSYTIAVSLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT	420	

|||||
Db 361 GPRSTTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
|
Db 421 M 421

RESULT 2

US-10-103-327-12
; Sequence 12, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-12

Query Match 100.0%; Score 2314; DB 2; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.4e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQLRNPHELHGCALALRFALVSWDIPGARALDNLGLARTPTMGWLHWFRCNLCQSEEP 60
Db 1 MQLRNPHELHGCALALRFALVSWDIPGARALDNLGLARTPTMGWLHWFRCNLCQSEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPORDSEGRLOADPQPFPHGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPORDSEGRLOADPQPFPHGIROL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGPGSGYGYDDAQTTFADWGVLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPGSGYGYDDAQTTFADWGVLLKFDGCGYCDLSLENL 180
QY 181 ADGKXMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIROYCNHWRNFPADIDDSWKSIX 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIROYCNHWRNFPADIDDSWKSIX 240
QY 241 SILDWTSTFQNERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFQNERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNDFEWEPLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNDFEWEPLSGLAWAVAMINRQETG 360
QY 361 GPRSTTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db 361 GPRSTTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
|
Db 421 M 421

RESULT 3

US-09-993-059-14
; Sequence 14, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-14

Query Match 100.0%; Score 2314; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQLRNPHELHGCALALRFALVSWDIPGARALDNLGLARTPTMGWLHWFRCNLCQSEEP 60
Db 1 MQLRNPHELHGCALALRFALVSWDIPGARALDNLGLARTPTMGWLHWFRCNLCQSEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPORDSEGRLOADPQPFPHGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPORDSEGRLOADPQPFPHGIROL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGPGSGYGYDDAQTTFADWGVLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPGSGYGYDDAQTTFADWGVLLKFDGCGYCDLSLENL 180
QY 181 ADGKXMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIROYCNHWRNFPADIDDSWKSIX 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIROYCNHWRNFPADIDDSWKSIX 240
QY 241 SILDWTSTFQNERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFQNERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNDFEWEPLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNDFEWEPLSGLAWAVAMINRQETG 360
QY 361 GPRSTTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db 361 GPRSTTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
|
Db 421 M 421

RESULT 4

US-10-103-327-14
; Sequence 14, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT

```
; ORGANISM: Homo sapiens
US-10-103-327-14

Query Match      100.0%; Score 2314; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-602-824A-2

Query Match      100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOLRNPELHLCALALRFLALVSDIPCARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLRNPELHLCALALRFLALVSDIPCARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGHKADAGEYELCIDDCWMAPOQDSGRLOADPORFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGHKADAGEYELCIDDCWMAPOQDSGRLOADPORFPHGIRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAOTFADWGYDLLKFDGCYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAOTFADWGYDLLKFDGCYCDLSLENL 180
Qy 181 ADGKHMSLALNRGSRIVSYCEWPLYMPPKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHMSLALNRGSRIVSYCEWPLYMPPKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSPQERIVDVAGPGGNDPDLVIGNFGLSWNOQVTOQALWAIMAAPLFMSNDL 300
Db 241 SILDWTSPQERIVDVAGPGGNDPDLVIGNFGLSWNOQVTOQALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOAKALLQDDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Qy 361 GPRSYTTIIVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLQLENT 420
Db 361 GPRSYTTIIVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLQLENT 420
Qy 421 M 421
Db 421 M 421

RESULT 5
US-07-602-824A-2
; Sequence 2, Application US/07602824A
; Patent No. 5356804
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; TITLE OF INVENTION: ACTIVE ALPHA-GALACTOSIDASE A
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/602,824A
; FILING DATE: 24-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-602-824A-2

Query Match      100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOLRNPELHLCALALRFLALVSDIPCARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLRNPELHLCALALRFLALVSDIPCARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGHKADAGEYELCIDDCWMAPOQDSGRLOADPORFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGHKADAGEYELCIDDCWMAPOQDSGRLOADPORFPHGIRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAOTFADWGYDLLKFDGCYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAOTFADWGYDLLKFDGCYCDLSLENL 180
Qy 181 ADGKHMSLALNRGSRIVSYCEWPLYMPPKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHMSLALNRGSRIVSYCEWPLYMPPKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSPQERIVDVAGPGGNDPDLVIGNFGLSWNOQVTOQALWAIMAAPLFMSNDL 300
Db 241 SILDWTSPQERIVDVAGPGGNDPDLVIGNFGLSWNOQVTOQALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOAKALLQDDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Qy 361 GPRSYTTIIVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLQLENT 420
Db 361 GPRSYTTIIVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLQLENT 420
Qy 421 M 421
Db 421 M 421

RESULT 6
US-07-602-608-3
; Sequence 3, Application US/07602608
; Patent No. 5382524
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Wang, Anne M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYL GALACTOSAMINIDASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/602,608
```

;
; FILING DATE: 24-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-07-602-608-3

Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMA PQDSEGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMA PQDSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGWVLLKFDGICYDSLENL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGWVLLKFDGICYDSLENL 180
QY 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKNYTEIROYCNHWRNPADIDDSWSKSIK 240
DB 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKNYTEIROYCNHWRNPADIDDSWSKSIK 240
QY 241 SILDWTFSNQERIVDVAGPGGWNDDMLVIGNFGLSMNQVVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTFSNQERIVDVAGPGGWNDDMLVIGNFGLSMNQVVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFEVWERPLSGLAWAVAMINRQIEIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFEVWERPLSGLAWAVAMINRQIEIG 360
QY 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 7
US-07-983-451-2
; Sequence 2, Application US/07983451
; Patent No. 5401650
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

;
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/983,451
; FILING DATE: 30-NOV-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 07/983,451
; REFERENCE/DOCKET NUMBER: 6923-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-983-451-2

Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMA PQDSEGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMA PQDSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGWVLLKFDGICYDSLENL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGWVLLKFDGICYDSLENL 180
QY 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKNYTEIROYCNHWRNPADIDDSWSKSIK 240
DB 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKNYTEIROYCNHWRNPADIDDSWSKSIK 240
QY 241 SILDWTFSNQERIVDVAGPGGWNDDMLVIGNFGLSMNQVVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTFSNQERIVDVAGPGGWNDDMLVIGNFGLSMNQVVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFEVWERPLSGLAWAVAMINRQIEIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFEVWERPLSGLAWAVAMINRQIEIG 360
QY 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 8
US-08-261-578-3
; Sequence 3, Application US/08261578
; Patent No. 5491075
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Wang, Anne M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY

TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYLGLACTOSAMINIDASE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,578
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/602,608
FILING DATE: 24-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6923-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-261-578-3

Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGMLHWRFMCNLDCQSEP 60
DB 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGMLHWRFMCNLDCQSEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120

QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCGYCDLSNLL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCGYCDLSNLL 180

QY 181 ADGKXMSLALNRTGRSIVYCEWPLYWMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKXMSLALNRTGRSIVYCEWPLYWMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

QY 241 SILDWTSTNQRIRVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLMSNDL 300
DB 241 SILDWTSTNQRIRVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLMSNDL 300

QY 301 RHISFOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQIG 360
DB 301 RHISFOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQIG 360

QY 361 GPRSYTIIAVASLGKGVACNPAFCITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
DB 361 GPRSYTIIAVASLGKGVACNPAFCITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420

QY 421 M 421
DB 421 M 421

RESULT 9
US-08-261-577-7
Sequence 7, Application US/08261577
Patent No. 5580757
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David F.
APPLICANT: Ioannou, Yiannis A.
TITLE OF INVENTION: Cloning and Expression of Biologically
TITLE OF INVENTION: Active alpha-Galactosidase A
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,577
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6923-042
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-261-577-7

Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGMLHWRFMCNLDCQSEP 60
DB 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGMLHWRFMCNLDCQSEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120

QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCGYCDLSNLL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCGYCDLSNLL 180

QY 181 ADGKXMSLALNRTGRSIVYCEWPLYWMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKXMSLALNRTGRSIVYCEWPLYWMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

QY 241 SILDWTSTNQRIRVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLMSNDL 300
DB 241 SILDWTSTNQRIRVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLMSNDL 300

QY 301 RHISFOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQIG 360
DB 301 RHISFOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQIG 360

QY 361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
Db |||||||
361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 10

US-08-261-577-9
; Sequence 9, Application US/08261577
; Patent No. 5580757
; GENERAL INFORMATION:
; APPLICANT: Deenick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,577
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-261-577-9

Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
Db |||||||
1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRQADPQPFPHGIRQL 120
Db |||||||
61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRQADPQPFPHGIRQL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGFPFGSGFYGYDDAQTADWGVLDLLKFDGCGYCDSENL 180
Db |||||||
121 ANYVHSGKLGIGIYADVGNKTCAGFPFGSGFYGYDDAQTADWGVLDLLKFDGCGYCDSENL 180
QY 181 ADGYKHMALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSITK 240
Db |||||||
181 ADGYKHMALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSITK 240
QY 241 SILDWTSTNQERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300

Db |||||||
241 SILDWTSTNQERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQETG 360
Db |||||||
301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
Db |||||||
361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 11

US-09-993-059-10
; Sequence 10, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-10

Query Match 100.0%; Score 2314; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
Db |||||||
1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRQADPQPFPHGIRQL 120
Db |||||||
61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRQADPQPFPHGIRQL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGFPFGSGFYGYDDAQTADWGVLDLLKFDGCGYCDSENL 180
Db |||||||
121 ANYVHSGKLGIGIYADVGNKTCAGFPFGSGFYGYDDAQTADWGVLDLLKFDGCGYCDSENL 180
QY 181 ADGYKHMALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSITK 240
Db |||||||
181 ADGYKHMALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSITK 240
QY 241 SILDWTSTNQERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db |||||||
241 SILDWTSTNQERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQETG 360
Db |||||||
301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
Db |||||||
361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
QY 421 M 421
Db 421 M 421


```
RESULT 12
US-10-103-327-10
; Sequence 10, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-10

Query Match      100.0%; Score 2314; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADWGVLLKFDGVCDSLENL 180
DB 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADWGVLLKFDGVCDSLENL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTIROYCNHWRNFADIDDSWSKSIK 240
DB 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTIROYCNHWRNFADIDDSWSKSIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFSNDL 300
DB 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQBIG 360
DB 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQBIG 360
QY 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 13
US-09-993-059-6
; Sequence 6, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
```

```
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-6

Query Match      100.0%; Score 2314; DB 2; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADWGVLLKFDGVCDSLENL 180
DB 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADWGVLLKFDGVCDSLENL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTIROYCNHWRNFADIDDSWSKSIK 240
DB 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTIROYCNHWRNFADIDDSWSKSIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFSNDL 300
DB 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQBIG 360
DB 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQBIG 360
QY 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 14
US-10-103-327-6
; Sequence 6, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-6

Query Match      100.0%; Score 2314; DB 2; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMCLDCQEEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMCLDCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRPFGHIGRL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRPFGHIGRL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGYYDIDAQTFADWGVLLKFDGCGYCDLNL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGYYDIDAQTFADWGVLLKFDGCGYCDLNL 180
QY 181 ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROCNHWRNFADIDDSWKSJK 240
Db 181 ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROCNHWRNFADIDDSWKSJK 240
QY 241 SILDWTSTFQNERIVDVAGPGWNDPDMVLVGNFGLSWNQVQTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFQNERIVDVAGPGWNDPDMVLVGNFGLSWNQVQTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWEPRLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWEPRLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 15

US-09-070-356-4
; Sequence 4, Application US/09070356
; Patent No. 6228631
; GENERAL INFORMATION:
; APPLICANT: Alex Zhu
; APPLICANT: Jack Goldstein
; TITLE OF INVENTION: Recombinant a-N-
; TITLE OF INVENTION: Acetylglactosaminidase
; TITLE OF INVENTION: Enzyme and cDNA Encoding
; TITLE OF INVENTION: Said Enzyme
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,356
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,248
; FILING DATE: March 26, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.
; REGISTRATION NUMBER: 34,894
; REFERENCE/DOCKET NUMBER: 63475/12
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 429
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: cDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: yes
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: human
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE: library
; POSITION IN GENOME: unknown
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY: human a-galactosidase
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS: Calhoun et al
; TITLE: Fabry Disease: Isolation of a cDNA
; TITLE: Clone Encoding Human a-Galactosidase A
; JOURNAL: Proceedings of the National Academy
; JOURNAL: of Science USA
; VOLUME: 82
; PAGES: 7364-7368
; DATE: 1985
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-09-070-356-4
Query Match 99.8%; Score 2310; DB 2; Length 429;
Best Local Similarity 99.0%; Pred. No. 3.9e-234;
Matches 417; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMCLDCQEEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMCLDCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRPFGHIGRL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRPFGHIGRL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGYYDIDAQTFADWGVLLKFDGCGYCDLNL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGYYDIDAQTFADWGVLLKFDGCGYCDLNL 180
QY 181 ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROCNHWRNFADIDDSWKSJK 240
Db 181 ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROCNHWRNFADIDDSWKSJK 240
QY 241 SILDWTSTFQNERIVDVAGPGWNDPDMVLVGNFGLSWNQVQTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFQNERIVDVAGPGWNDPDMVLVGNFGLSWNQVQTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWEPRLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWEPRLSGLAWAVAMINRQETG 360

Qy	361	GPRSYTIAVASLGKGVACNPACFITQLLPVKXKLGFEWTSRLRSHINPTGTVLLQLENT	420
Db	361	GPRSYTIAVASLGKGVACNPACFITQLLPVKXKLGFEWTSRLRSHINPTGTVLLQLENT	420
Qy	421	M	421
Db	421	M	421

Search completed: December 24, 2005, 09:12:00
Job time : 45 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 09:07:49 ; Search time 40 Seconds
(without alignments)
1012.681 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQRNPEHLGICALAFLA.....RLRSHINPTGVLLQLENTM 421
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 80:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2314	100.0	429	1 GBHUA	alpha-galactosidase
2	1811	78.3	419	2 JC4522	alpha-galactosidase
3	1051	45.4	405	2 S45522	alpha-N-acetylgalactosaminidase
4	952.5	41.2	411	2 A35485	alpha-N-acetylgalactosaminidase
5	930	40.2	358	2 A33265	alpha-N-acetylgalactosaminidase
6	862	37.3	451	2 T24018	hypothetical protein
7	716.5	31.0	434	2 T47748	alpha-galactosidase
8	712	30.8	378	2 T50781	alpha-galactosidase
9	687.5	29.7	422	2 T06388	alpha-galactosidase
10	682	29.5	425	2 T10860	alpha-galactosidase
11	681.5	29.5	411	2 S07472	alpha-galactosidase
12	676	29.2	680	2 T36472	probable secreted
13	609	26.3	436	2 T39118	probable alpha-galactosidase
14	598	25.8	396	2 JC5558	alpha-galactosidase
15	557	24.1	469	2 S45453	alpha-galactosidase
16	543.5	23.5	545	2 S23582	alpha-galactosidase
17	521	22.5	471	2 S50312	alpha-galactosidase
18	519	22.4	471	2 S50311	alpha-galactosidase
19	516	22.3	444	2 T74221	alpha-galactosidase
20	514	22.2	471	2 JQ1021	alpha-galactosidase
21	510	22.0	471	2 S50310	alpha-galactosidase
22	502	21.7	471	1 GBBYAG	hypothetical protein
23	300.5	13.0	204	2 T04423	probable alpha-galactosidase
24	282.5	12.2	432	2 F83883	alpha-galactosidase
25	198	8.6	159	2 T04422	alpha-galactosidase
26	180.5	7.8	624	2 S74222	alpha-galactosidase
27	129	5.6	641	2 A55549	glucan 1,6-alpha-D-glucanase
28	125	5.4	348	2 A70311	hypothetical protein
29	116	5.0	4199	2 S76412	hypothetical protein

RESULT 1

GBHUA

alpha-galactosidase (EC 3.2.1.22) A precursor - human

N:Alternate names: alpha-D-galactoside galactohydrolase; melibiase

C:Species: Homo sapiens (man)

C:Date: 30-Jun-1987, #sequence revision 27-Oct-1995 #text_change 09-Jul-2004

C:Accession: S04081; A29608; A30214; S14879; A00896; B00896; I37140

R:Kornreich, R.; Desnick, R.J.; Bishop, D.F.

Nucleic Acids Res. 17, 3301-3302, 1989

A:Title: Nucleotide sequence of the human alpha-galactosidase A gene.

A:Reference number: S04081; MUID:89263745; PMID:2542896

A:Accession: S04081

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-429 <KOR>

A:Cross-references: UNIPROT:P06280; UNIPARC:UPI00000033A30; EMBL:X14448; NID:g31755; PIDN:Q14448

R:Quinn, M.; Hantzopoulos, P.; Fidanza, V.; Calhoun, D.H.

Gene 58, 177-188, 1987

A:Title: A genomic clone containing the promoter for the gene encoding the human lysosomal

A:Reference number: A29608; MUID:88112869; PMID:2892762

A:Accession: A29608

A:Molecule type: DNA

A:Residues: 1-64 <QUI>

A:Cross-references: UNIPARC:UPI000016A96A; GB:M18242; NID:g182944; PIDN:AAA52514.1; PID:Q182944

R:Bishop, D.F.; Kornreich, R.; Desnick, R.J.

Proc. Natl. Acad. Sci. U.S.A. 85, 3903-3907, 1988

A:Title: Structural organization of the human alpha-galactosidase A gene: further evidence

A:Reference number: A30214; MUID:88234528; PMID:2836863

A:Accession: A30214

A:Molecule type: DNA

A:Residues: 1-64 <BIS>

A:Cross-references: UNIPARC:UPI000016A96A; EMBL:M20317; EMBL:J03249

R:Koide, T.; Ishiura, M.; Iwai, K.; Inoue, M.; Kaneda, Y.; Okada, Y.; Uchida, T.

FEBS Lett. 259, 353-356, 1990

A:Title: A case of Fabry's disease in a patient with no alpha-galactosidase A activity

A:Reference number: S14879; MUID:90092580; PMID:2152885

A:Accession: S14879

A:Molecule type: mRNA

A:Residues: 1, 'K', 'J', '3', '39', 'S', '41-429 <KOI>

A:Cross-references: UNIPARC:UPI000011E285; EMBL:X16989

A:Experimental source: Fabry's disease patient

R:Bishop, D.F.; Calhoun, D.H.; Bernstein, H.S.

Proc. Natl. Acad. Sci. U.S.A. 83, 4859-4863, 1986

A:Title: Human alpha-galactosidase A: nucleotide sequence of a cDNA clone encoding the

A:Reference number: A00896; MUID:86259694; PMID:3014515

A:Accession: A00896

A:Molecule type: mRNA

A:Residues: 27-429 <B12>

A:Cross-references: UNIPARC:UPI0000000358; GB:M13571; NID:g178245; PIDN:AAA51676.1; PID:Q178245

A:Experimental source: lung

A:Accession: B00896

A:Molecule type: protein

A;Residues: 32-55,'S',57-58,'R',60-65,'S',67-68;228-232,'N',234-237,'A',298-326;'L',334-
A;Cross-references: UNIPARC:UPI00001729C6; UNIPARC:UPI00001729C7; UNIPARC:UPI00001729C8;
R;Tsuiji, S.; Martin, B.M.; Kaslow, D.C.; Migeon, B.R.; Choudary, P.V.; Stubblefield, B.K
Eur. J. Biochem. 165, 275-280, 1987
A;Title: Signal sequence and DNA-mediated expression of human lysosomal alpha-galactosid
A;Reference number: 137140; PMID:87246603; PMID:3036505
A;Accession: J37140
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-429 <RES>
A;Cross-references: UNIPARC:UPI0000033A30; EMBL:X05790; NID:g28535; PIDN:CAA29232.1; PID
C;Genetics:
A;Gene: GDB:GLA
A;Cross-references: GDB:119272; OMIM:301500
A;Map position: Xq21.3-Xq22
A;Introns: 65/2; 123/3; 183/1; 213/3; 267/3; 333/3
C;Superfamily: alpha-galactosidase
C;Keywords: Fabry disease; glycolipid metabolism; glycoprotein; glycosidase; hydrolase;
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-429/Product: alpha-galactosidase A #status predicted <MAT>
F;139,192,215,408/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.7e-190;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLNPEHLGALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
DB 1 MOLNPEHLGALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYVLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYVLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGCGYCDLSNLL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGCGYCDLSNLL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNFPADIDDSWSKSIK 240
DB 181 ADGYKHSALNRTGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNFPADIDDSWSKSIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTOMALWAIMAAPLFWNDL 300
DB 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTOMALWAIMAAPLFWNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYLRQGNFVWRPLSLGLAWAVAMINRQBIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYLRQGNFVWRPLSLGLAWAVAMINRQBIG 360
QY 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421
RESULT 2
JC4522
alpha-galactosidase (EC 3.2.1.22) A precursor - mouse
N;Alternate names: alpha-D-galactoside galactohydrolase
C;Species: Mus musculus (house mouse)
C;Date: 08-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C;Accession: JC4522
R;Ohshima, T.; Murray, G.J.; Nagle, J.W.; Quirk, J.M.; Kraus, M.H.; Barton, N.W.; Brady,
Gene 166, 277-280, 1995
A;Title: Structural organization and expression of the mouse gene encoding alpha-galacto
A;Reference number: JC4522; PMID:96125203; PMID:8541375
A;Accession: JC4522
A;Molecule type: mRNA
A;Residues: 1-419 <OHS>

A;Cross-references: UNIPROT:P51569; UNIPARC:UPI00000018CA; GB:U34071; NID:gl141787; PIDN:
A;Experimental source: kidney, C57BL
C;Comment: This enzyme is a lysosomal enzyme that hydrolyses the alpha-D-galactosyl resi
C;Genetics:
A;Gene: alpha Gal
A;Introns: 65/2; 123/3; 183/1; 213/3; 267/3; 333/3
C;Superfamily: alpha-galactosidase
C;Keywords: glycoprotein; glycosidase; hydrolase; lysosome
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-419/Product: alpha-galactosidase A #status predicted <MAT>
F;139,192,215,408/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 78.3%; Score 1811; DB 2; Length 419;
Best Local Similarity 78.2%; Pred. No. 2.3e-147;
Matches 327; Conservative 41; Mismatches 50; Indels 0; Gaps 0;
QY 1 MOLNPEHLGALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
DB 1 MKLLSRDTRLVCELALCPALVFWLSILGVRLDNGLARTPTMGWLHWRFCNLDQCEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYVLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
DB 61 DACISEQLFMQAEMLVSDGWRDAGYDYLICIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGCGYCDLSNLL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGCGYCDLSNLL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNFPADIDDSWSKSIK 240
DB 181 ENGYKYMALNRTGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNFPADIDDSWSKSIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTOMALWAIMAAPLFWNDL 300
DB 241 NILSWTVVYQKEIVEVAGPGGWNDDMLVIGNFGLSWNQVQTOMALWAIMAAPLFWNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYLRQGNFVWRPLSLGLAWAVAMINRQBIG 360
DB 301 RQISSQAKALLQNDVIAINQDPLGKQGYCFRKNHIEVWRPLSLGLAWAVAMINRQBIG 360
QY 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLE 419
DB 361 GPCPTYIQISSIGRLACNPCCIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLE 419
RESULT 3
S45522
alpha-N-acetylglucosaminidase - chicken
C;Species: Gallus gallus (chicken)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S45522; S43413
R;Davis, M.O.; Hata, J.; Smith, D.; Walker, J.C.
submitted to the EMBL Data Library, December 1993
A;Reference number: S45522
A;Accession: S45522
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-405 <DAV>
A;Cross-references: UNIPROT:Q90744; UNIPARC:UPI000000FC7C2; EMBL:L18754; NID:g435036; PID
R;Davis, M.O.; Hata, D.J.; Smith, D.; Walker, J.C.
Biochim. Biophys. Acta 1216, 296-298, 1993
A;Title: Cloning and sequence of a chicken alpha-N-acetylglucosaminidase gene.
A;Reference number: S43413; PMID:94060104; PMID:8241271
A;Accession: S43413
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-161,'A',163-405 <DA2>
A;Cross-references: UNIPARC:UPI0000146A75; GB:L18754
C;Superfamily: alpha-galactosidase
Query Match 45.4%; Score 1051; DB 2; Length 405;
Best Local Similarity 53.0%; Pred. No. 3.2e-82;
Matches 206; Conservative 58; Mismatches 117; Indels 8; Gaps 6;

Qy 32 LDNGLARTPTMGWLHWFRCMCLDCQEEPDSCISEKLFMEAMVSEGWKDGAYEYLICI 91
Db 1 LENGGLARTPTMGWLHWFRCMCLDCQEEPDSCISEKLFMEAMVSEGWKDGAYEYLICI 60
Qy 92 DDCWMAQORDESEGLQADPQPFPHGIRQLANYVHSKGLGIYADVGNKTCAGPFG-SFG 150
Db 61 DDCWAAQORDESEGLQADPQPFPHGIRQLANYVHSKGLGIYADVGNKTCAGPFG-SFG 120
Qy 151 YVDIDAQTFADGWGVDLLKFDGCGYCDLSLENLADGKYHMSLALNRTGRSIVYSCWPLYMWP 210
Db 121 RVEQDAQTFADGWGVDLLKFDGCGYCDLSLENLADGKYHMSLALNRTGRSIVYSCWPLYMWP 179
Qy 211 F-QKPNYTEIRQYCNHWRNFADIDDSWKSISILDTWTSFNQERIVDVAGPGWNPDMVLV 269
Db 180 LPPKVNYYTLGECICNLWRNYDDIQDSWMSVLSIVDWFPTNQDVLQPFAGPGHWNPDMLI 239
Qy 270 IGNEGLSNQVQTMALWMAAPLFWNSDLRHISPOKALLQKDVTAIINQDPLGKQGY 329
Db 240 IGNEGLSYEQRSQNALWTMAAPLMTDLRTISPSAKTILONRLMTIQINQDPLGIQGR 299
Qy 330 Q-LRGDNFVYWERPLSLGLAWAVAMINRQIEGGPRSYTIAVASLGKGVACNPACFITQLL 388
Db 300 RIIEGSHIEVFLRPLSQASALVFFSRR-TDMFRYTTSLAKLGFPMG---AAYEVQDV 355
Qy 389 PVKRLGFGYEWTSRLRSHINPTGTVLLQL 417
Db 356 YSGKIISGLKTGDNFTVIINPSGVVMWYL 384

RESULT 4
A35485
alpha-N-acetylgalactosaminidase (EC 3.2.1.49) precursor, splice form 2 - human
N:Alternate names: alpha-galactosidase B
C:Species: Homo sapiens (man)
C:Date: 31-Aug-1990 #sequence_revision 31-Dec-1995 #text_change 09-Jul-2004
C:Accession: A36530; A35485; A37082; I39416
R:Wang, A.M.; Bishop, D.F.; Desnick, R.J.
J. Biol. Chem. 265, 21859-21866, 1990
A:Title: Human alpha-N-acetylgalactosaminidase-molecular cloning, nucleotide sequence, and ancestral gene.
A:Reference number: A36530; MUID:91072392; PMID:2174888
A:Accession: A36530
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-411 <YAM>
A:Cross-references: UNIPROT:P17050; UNIPARC:UPI000012FD03; GB:M62783; NID:g178247; PIDN:
R:Yamauchi, T.; Hiraiwa, M.; Kobayashi, H.; Uda, Y.; Miyatake, T.; Tsuji, S.
Biochem. Biophys. Res. Commun. 170, 231-237, 1990
A:Title: Molecular cloning of two species of cDNAs for human alpha-N-acetylgalactosaminidase.
A:Reference number: A35485; MUID:90321233; PMID:2372288
A:Accession: A35485
A:Molecule type: mRNA
A:Residues: 1-411 <YAM>
A:Cross-references: UNIPARC:UPI000012FD03; GB:M38083; NID:g189054; PIDN:AAA36351.1; PID:
R:Yamauchi, T.; Hiraiwa, M.; Kobayashi, H.; Uda, Y.; Miyatake, T.; Tsuji, S.
Biochem. Biophys. Res. Commun. 170, 231-237, 1990
A:Title: Photolabeling of the alpha-neuraminidase/beta-galactosidase complex from human
A:Reference number: A37082; MUID:91076941; PMID:2256909
A:Accession: A37082
A:Molecule type: protein
A:Residues: 18-23, 'N', 25-29, 'X', 31-32, 'X', 34-37 <WAB>
A:Cross-references: UNIPARC:UPI0000175B1D
A:Experimental source: placenta
R:Wang, A.M.; Desnick, R.J.
Genomics 10, 133-142, 1991
A:Title: Structural organization and complete sequence of the human alpha-N-acetylgalactosaminidase gene.
A:Reference number: I39416; MUID:91257820; PMID:1646157
A:Accession: I39416
A:Status: preliminary; translation not shown

A:Molecule type: DNA
A:Residues: 1-391 <RES>
A:Cross-references: UNIPARC:UPI0000175B1E; GB:M59199; NID:g1513066; PIDN:AA06718.1; PID:
A:Note: sequence extracted from NCBI backbone and corrected to correspond with the trans
C:Genetics:
A:Gene: GDB:NAGA
A:Cross-references: GDB:119445; OMIM:104170
A:Map position: 22q11-22q11
A:Introns: 6/1; 51/2; 108/3; 168/1; 199/3; 253/3; 319/3; 367/3
C:Superfamily: alpha-galactosidase
C:Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase; lysosome
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-411/Product: alpha-N-acetylgalactosaminidase, splice form 2 #status predicted <MAT>
F:124,177,201,359,385,391/Binding site: carbohydrate (Aen) (covalent) #status predicted
F:124,177,201,359,385,391/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 41.2%; Score 952.5; DB 2; Length 411;
Best Local Similarity 50.0%; Pred. No. 9e-74;
Matches 198; Conservative 55; Mismatches 120; Indels 23; Gaps 9;

Qy 32 LDNGLARTPTMGWLHWFRCMCLDCQEEPDSCISEKLFMEAMVSEGWKDGAYEYLICI 91
Db 18 LDNGLLQTPPMGWLAWERFRCNINCDKNCISEQLFMEAMADMAQDGRMDMGTYLNI 77
Qy 92 DDCWMAQORDESEGLQADPQPFPHGIRQLANYVHSKGLGIYADVGNKTCAGPFG-SFG 150
Db 78 DDCWIG-GRDASGRMLPDPKRPFGHIGIPFLADYVHSLGKGLGIYADMGNTFCMGYPTGLD 136
Qy 151 YVDIDAQTFADGWGVDLLKFDGCGYCDLSLENLADGKYHMSLALNRTGRSIVYSCWPLYMWP 210
Db 137 KVQDAQTFADGWGVDLLKFDGCGYCDLSLENLADGKYHMSLALNRTGRSIVYSCWPLYMWP 195
Qy 211 F-QKPNYTEIRQYCNHWRNFADIDDSWKSISILDTWTSFNQERIVDVAGPGWNPDMVLV 269
Db 196 LPPRVNYSLLADICNLWRNYDDIQDSWMSVLSILNWFVEHQDILQPVAGPGHWNPDMLL 255
Qy 270 IGNEGLSNQVQTMALWMAAPLFWNSDLRHISPOKALLQKDVTAIINQDPLGKQGY 329
Db 256 IGNEGLSLEQSAQWALMTVLAAPLMTDLRTISPSAKTILONRLMTIQINQDPLGIQGR 315
Qy 330 Q-LRGDNFVYWERPLSLGLAWAVAMIN-ROIEGGPRSYTIAVASLGKGVACNPACFITQL 387
Db 316 RIIEGSHIEVFLRPLSQASALVFFSRR-TDMFRYTTSLAKLGFPMG---FTGSV 364
Qy 388 LPVXKRLGFGYEWTSRLRSH-----INPTGTVLLQL 417
Db 365 IYEAQDVYSGDIISGLRDETNFTVIINPSGVVMWYL 400

RESULT 5
A33265
alpha-N-acetylgalactosaminidase (EC 3.2.1.49) precursor, splice form 1 - human
N:Alternate names: alpha-galactosidase B
C:Species: Homo sapiens (man)
C:Date: 21-Feb-1990 #sequence_revision 21-Feb-1990 #text_change 09-Jul-2004
C:Accession: A33265
R:Tsuji, S.; Yamauchi, T.; Hiraiwa, M.; Isobe, T.; Okuyama, T.; Sakimura, K.; Takahashi,
Biochem. Biophys. Res. Commun. 163, 1498-1504, 1989
A:Title: Molecular cloning of a full-length cDNA for human alpha-N-acetylgalactosaminidase.
A:Reference number: A33265; MUID:89392067; PMID:2551294
A:Accession: A33265
A:Molecule type: mRNA
A:Residues: 1-358 <TSU>
A:Cross-references: UNIPROT:P17050; UNIPARC:UPI000016ADB4; GB:M29276; NID:g189052; PIDN:
A:Experimental source: clone pCD-HS1204
C:Genetics:
A:Gene: GDB:NAGA
A:Cross-references: GDB:119445; OMIM:104170
A:Map position: 22q11-22q11
A:Introns: 6/1; 51/2; 108/3; 168/1; 199/3; 253/3; 319/3
C:Superfamily: alpha-galactosidase
C:Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase; lysosome
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-358/Product: alpha-N-acetylgalactosaminidase, splice form 1 #status predicted <MAT>

R.Zhu, A.; Goldstein, J.
Gene 140, 227-231, 1994
A>Title: Cloning and functional expression of a cDNA encoding coffee bean alpha-galactosidase
A;Reference number: Z25235; MUID:94193002; PMID:8144030
A;Accession: T50781
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-378 <ZHU>
A;Cross-references: UNIPROT:Q42656; UNIPARC:UPI00001256A8; EMBL:L27992; PIDN:AAA33022.1
C;Superfamily: alpha-galactosidase
C;Keywords: glycosidase; hydrolase

Query Match 30.8%; Score 712; DB 2; Length 378;
Best Local Similarity 42.5%; Pred. No. 3.4e-53;
Matches 152; Conservative 51; Mismatches 113; Indels 42; Gaps 8;
QY 32 LDNGLARTPTMGWLHWRFCMNLDCQEPDSCISEKLFMEAEMLVSEGWKADAGYEYLCI 91
DB 16 LANGLGLTPPMGWSNNHFRCLD-----EKLIREADAMVSKGLAALGYKYINL 65
QY 92 DDCWAPQDSEGRGLQADPQRPFGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSFG 150
DB 66 DDCWAEALNDSQNLVPRGSTPFGIKALADYVHSGKLGKIYSDAGTQTCSTKTPGSLG 125
QY 151 YYDIDAQTFADGWDLKFDGCGYCDSELENLADGYKHMSLALNRTORSIVYS-CEWPLYMW 209
DB 126 HEEQDAKTFASWGIDYLYDNCNNNI--SPKERYPIMS KALINSGRSIFFSCLCEWG---- 180
QY 210 PFQKNYETIRQYCNHWNFNADIDDSWKSIIKILD----WTSFNQERIVDVAGPGGWNNDP 265
DB 181 --SEDPATWAKSVGNSWRTTGDIDDSWSSMTSRADNDKWASY-----AGPGGWNNDP 230
QY 266 DMLVIGNFGLSWNOQVQTMALWMAAPLFMSNDRHISPOAKALLODKVDIAINQDPLG 325
DB 231 DMLVGVNGGMMTTEYSRHSFISWALAKAPLLIGCDIRSDMGATFQLLSNAEVIANNQDKLG 290
QY 326 KQGYQLRGQDNFVWERPLSGLAWAVAMINR-----QETGGPRSYTIAVASL 372
DB 291 VQGNKVKTYGDLEWVAGPLSGKRVAVALNWRGSSATITAYMSDVGLPSTAVVNAARDL 348

RESULT 9
T06388
alpha-galactosidase (EC 3.2.1.22) - soybean
C;Species: Glycine max (soybean)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
C;Accession: T06388
R;Davis, M.O.; Walker, J.C.; Smith, D.
submitted to the EMBL Data Library, August 1994
A;Description: Cloning and expression of a soybean alpha galactosidase gene.
A;Reference number: Z15645
A;Accession: T06388
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-422 <DAV>
A;Cross-references: UNIPROT:Q39811; UNIPARC:UPI00000A7200; EMBL:U12926; PID:G927574; PID
A;Experimental source: strain williams
C;Function:
A;Description: catalyzes hydrolysis of melibiose into galactose and glucose
C;Superfamily: alpha-galactosidase
C;Keywords: glycosidase; hydrolase

Query Match 29.7%; Score 687.5; DB 2; Length 422;
Best Local Similarity 43.2%; Pred. No. 4.9e-51;
Matches 143; Conservative 45; Mismatches 112; Indels 31; Gaps 7;
QY 32 LDNGLARTPTMGWLHWRFCMNLDCQEPDSCISEKLFMEAEMLVSEGWKADAGYEYLCI 91
DB 60 LDNGLHTTPPMGWSNNHFRCLD-----IKEDLIREADAMVSTGLAALGYQYINI 109
QY 92 DDCWAPQDSEGRGLQADPQRPFGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSFG 150
DB 110 DDCWELNDRSGNLVPRKASTPFGMKALADYVHKNGLKGIYSDAGNQTCSKTTPGSLG 169

QY 151 YYDIDAQTFADGWDLKFDGCGYCDSELENLADGYKHMSLALNRTORSIVYS-CEWPLYMW 209
DB 170 HEEQDAKTFASWGIDYLYDNCNNNI--SPKERYPMPMSEALANTGRPIFFSLCEWG----- 224
QY 210 PFQKNYETIRQYCNHWNFNADIDDSWKSIIKILD----WTSFNQERIVDVAGPGGWNNDP 265
DB 225 --SEDPATWAKSVGNSWRTTGDIDQKWSMISRADLNDKWASY-----AGPGGWNNDP 274
QY 266 DMLVIGNFGLSWNOQVQTMALWMAAPLFMSNDRHISPOAKALLODKVDIAINQDPLG 325
DB 275 DMLVGVNGGMMTTEYSRAHFSIWSLAKAPLLIGCDIRALDATTKELLSKEVIANNQDKLG 334
QY 326 KQGYQLRGQDNFVWERPLSGLAWAVAMINR 356
DB 335 VQGKKVKSTNDLEWVAGPLSNKVKAVILWNR 365

RESULT 10
T10860
alpha-galactosidase (EC 3.2.1.22) - kidney bean
C;Species: Phaseolus vulgaris (kidney bean)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T10860
R;Davis, M.O.; Walker, J.C.; Smith, D.
submitted to the EMBL Data Library, August 1994
A;Description: Cloning and expression of a pinto bean alpha galactosidase gene.
A;Reference number: Z17189
A;Accession: T10860
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-425 <DAV>
A;Cross-references: UNIPROT:Q41100; UNIPARC:UPI00000A0F13; EMBL:U12927; NID:G927576; PID
C;Function:
A;Description: catalyzes hydrolysis of melibiose into galactose and glucose
C;Superfamily: alpha-galactosidase
C;Keywords: glycosidase; hydrolase

Query Match 29.5%; Score 682; DB 2; Length 425;
Best Local Similarity 41.5%; Pred. No. 1.5e-50;
Matches 144; Conservative 54; Mismatches 113; Indels 36; Gaps 9;
QY 19 LALVSWDIPGARAL-DNGLARTPTMGWLHWRFCMNLDCQEPDSCISEKLFMEAEMLV 77
DB 49 MMMSREVDHRRNLVGNGLGQTTPMGWSNNHFSN-----INEDLIREADAMV 98
QY 78 SEGKQDAGYEYLCIDDCWMAQDSEGRGLQADPQRPFGIRQLANYVHSGKLGKIYADV 137
DB 99 STGLAALGYQYINIDDCWELNDRSGNLVPRKASTPFGMKALADYVHKNGLKGIYSDA 158
QY 138 GNKTCA-GPFGSFGYDIDAQTFADGWDLKFDGCGYCDSELENLA--DGYKHMSLALNRT 194
DB 159 GTQCTSKTTPGSLGHEEQDAKTFASWGIDYLYDNC---ENKNISPKERYPPMPSKALANS 215
QY 195 GRSIVYS-CEWPLYMWPPQKNYETIRQYCNHWNFNADIDDSWKSIIKILD----WTSFN 249
DB 216 GRPIFFSLCEWG-----SEDPATWAKSVGNSWRTTGDIEDKWSMISRADLNDWASY- 268
QY 250 QERIVDVAGPGGWNNDPDMLVIGNFGLSWNOQVQTMALWMAAPLFMSNDRHISPOAKA 309
DB 269 -----AGPGGWNNDPDMLEVGNGGMMTTEYSRAHFSIWSLAKAPLLIGCDIRALDVTYKE 321
QY 310 LLQDKVDIAINQDPLGKQGYQLRGQDNFVWERPLSGLAWAVAMINR 356
DB 322 LLSNEEVIANNQDKLGQGVKKVKNNDLEWVAGPLSNKVKAVILWNR 368

RESULT 11
S07472
alpha-galactosidase (EC 3.2.1.22) precursor - guar
C;Species: Cyanopsis tetragonoloba (guar, cluster bean)
C;Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 09-Jul-2004
C;Accession: S07472

Db 341 YIELPSGRLSNDDWVAVLN 360

RESULT 14

JCS558
alpha-galactosidase (EC 3.2.1.22) II precursor - Mortierella vinacea
N:Alternate names: melibiase
C:Species: Mortierella vinacea
C>Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 09-Jul-2004
C:Accession: JCS558; PC4479
R:Shibuya, H.; Kobayashi, H.; Sato, T.; Kim, W.S.; Yoshida, S.; Kaneko, S.; Kasamo, K.; Biosci. Biotechnol. Biochem. 61, 592-598, 1997
A:Title: Purification, characterization, and cDNA cloning of a novel alpha-galactosidase
A:Reference number: JCS558; MUID:97290877; PMID:9145516
A:Accession: JCS558
A:Molecule type: mRNA
A:Residues: 1-396 <SH1>
A:Cross-references: UNIPROT:Q93816; UNIPARC:UPI000006AE03; GB:AB018691; NID:G3777480; PI
A:Molecule type: protein
A:Residues: 21-39; 178-187 <SH2>
A:Cross-references: UNIPARC:UPI0000175B1F; UNIPARC:UPI0000175B20
A:Note: the authors translated the codon TCC for residue 112 as Leu, ACC for residue 122
e 236 as Gln
A:Note: the authors translated the codon ATG for residue 268 as Asn, TCC for residue 310
C:Comment: This enzyme liberates both side-chain and terminal alpha-galactosyl residues
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-396/Product: alpha-galactosidase II #status predicted <MAT>

Query Match 25.8%; Score 598; DB 2; Length 396;
Best Local Similarity 37.4%; Pred. No. 2.1e-43;
Matches 126; Conservative 57; Mismatches 116; Indels 38; Gaps 9;

QY 32 LNLGARTPTMGWLHWFRCNLDCEEPDSCISEKLFMEAEMLMVESECKDAGVEYLCT 91
Db 22 IDPSLAKTQMGNSNKKYQCN-----VNETVIINTANAVSSGLKDLGVHYINI 71

QY 92 DDCWMAPORD-SEGRLOADPQRPFGHGIROLANYVSKGLKGIYADVGNKTCAGPFGSFG 150
Db 72 DDCWLSLHQDNTTQTAPDTPFPNGISGVASKVHALGLKGIYSDAGTNTCAGYFGSYG 131

QY 151 YVDIDQTFADWGVLLKPFDCYCDSEML-----ADGYKMSLALNTRGRSIVYS 201
Db 132 YEADAQAQFSDWGVLYKYDNC-----NNLGLAGNATISSKRYKRGMDALKNVSRPIFFS 186

QY 202 -CEWPL-YMPPQKPNYTEIRQYCNHWRNFPADIDDSWKSISILDTWTSFNQERIVDVAGP 259
Db 187 LCSWGTDDVDWGRSTGGQ-----SWRMSGDISDNWSSVSI---TQAVPIANISAP 236

QY 260 GGWNPDPMLVIG-NFGLSNQVQTOMALWAIAPLFMSNDRHISPOAKALLQDKDVIA 318
Db 237 GGWNDMDLVEVGHDDMTITETSHPSIWAAMKSPILINDITNMTNDIKNIITNEVIA 296

QY 319 INQDPLGKQGYQLRQGDNFVEMRPLSLGLAWAVAMIN 355
Db 297 ISQDSLGSVQQRSMKGNTOQLFAGPLSKNGYVSLFLN 333

RESULT 15

S45453
alpha-galactosidase (EC 3.2.1.22) MEL precursor - yeast (Zygosaccharomyces cidri)
N:Alternate names: MEL protein
C:Species: Zygosaccharomyces cidri
C>Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
C:Accession: S45453
R:Turakainen, H.; Hankaanpaae, M.; Korhola, M.; Aho, S.
Yeast 10, 733-745, 1994
A:Title: Characterization of MEL genes in the genus Zygosaccharomyces.
A:Reference number: S45453; MUID:95066377; PMID:7975892
A:Accession: S45453
A:Molecule type: DNA

A:Residues: 1-469 <TUR>
A:Cross-references: UNIPROT:Q99172; UNIPARC:UPI00000696F1; EMBL:L24957; NID:9538518; PI
C:Genetics:
A:Gene: MEL
C:Superfamily: alpha-galactosidase
C:Keywords: glycoprotein; glycosidase; hydrolase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-469/Product: alpha-galactosidase MEL #status predicted <MAT>

Query Match 24.1%; Score 557; DB 2; Length 469;
Best Local Similarity 35.3%; Pred. No. 8.7e-40;
Matches 134; Conservative 52; Mismatches 128; Indels 66; Gaps 12;

QY 18 FLALV--SWDIPGARALDNLARTPTMGWLHWRFCNLDCEEPDSCISEKLFMEAEML 75
Db 5 FFALFFSSTDVLAASPSVNLGLTQMGMNDNNSFGCS-----VKEELLGLTAEK 54

QY 76 MYSEGKADAGVEYLCIDDCWMAPORDSEGRLOADPQRPFGHGIROLANYVSKGLKGIYA 135
Db 55 IVKLGKLDLGYNYIILDDCW-SSGRSSNGSLDADSKFPFGMKYVAEQJHNSQLKFGMYS 113

QY 136 DVGNTKTCAGPFGSFGYDIDDAQTTFADWGVLLKFGDCY-----CDSLENLADGYKHMSLAL 191
Db 114 SAGEVTCAGYAGSLGYEDMDATFASWDVLYKYDNCYNKGFEGTPEISYKRYKMSDAL 173

QY 192 NRTGRSIVYS-CEWP---LYMPPQKPNYTEIRQYCNHWRNFPADI----- 232
Db 174 NKTGRPIFYSLCNWQDLTFYWG-----SAISNSWRMSGDVVYPQFDRPDSRCPGSG 224

QY 233 ---DDSWK-----SISILDWTSFNQERIVDVAGPFGWNPDMIVGNFGLSNQVQTOMA 285
Db 225 DEYDCSYFGFHCISIMNLNKAAPMGON-----AAPGWNLDMLLEVGVGNMSDSEVAHFS 280

QY 286 LWAIMAAPLFMSNDRHISPOAKALLQDKDVIAINQDPLGKQ-----YQLRQGDNF--- 337
Db 281 MWAIYKSPILIGADIDDLKDSLSVSNPAVIAINQDVLGTPTATRWKTHVSDKQOYGE 340

QY 338 --EVMERPLSLGLAWAVAMIN 355
Db 341 ETQLWSGPLDNGDHVVALLN 360

Search completed: December 24, 2005, 09:12:46
Job time : 42 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 04:08:03 ; Search time 6366 Seconds
(without alignments)
11304.400 Million cell updates/sec

Title: US-10-602-219-11
Perfect score: 1266
Sequence: 1 atgcagctgaggaaccacaga.....agctagaacacaaatgtaa 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues ;
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_in.*
- 3: gb_env.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pr.*
- 9: gb_ro.*
- 10: gb_sts.*
- 11: gb_ev.*
- 12: gb_un.*
- 13: gb_vl.*
- 14: gb_htg.*
- 15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	100.0	1266	6	AR653696
2	1266	100.0	1266	6	AR654964
3	1264	99.8	1284	6	AR653697
4	1264	99.8	1284	6	AR654965
5	1263.4	99.8	1278	6	AR653694
6	1263.4	99.8	1278	6	AR654962
7	1263.4	99.8	1290	6	AR653692
8	1263.4	99.8	1296	6	AR654960
9	1263.4	99.8	1296	6	AR653695
10	1263.4	99.8	1296	6	AR654963
11	1263.4	99.8	1308	6	AR653693
12	1263.4	99.8	1308	6	AR654961
13	1261.8	99.7	1290	6	CS135670
14	1261.8	99.7	1290	11	BT006864
15	1261.8	99.7	1290	11	AY889631
16	1261.8	99.7	1290	11	AY889632
17	1261.8	99.7	1290	11	AY889209
18	1261.8	99.7	1290	11	BT007835

19	1261.8	99.7	1306	6	BD182986	BD182986	Pharmaceu
20	1261.8	99.7	1306	6	BD185988	BD185988	Pharmaceu
21	1261.8	99.7	1312	8	BC002689	BC002689	Homo sapi
22	1261.8	99.7	1319	6	AX840234	AX840234	Sequence
23	1261.8	99.7	1319	8	HSAGALAR	HSAGALAR	X05790 Human mRNA
24	1261.8	99.7	1343	6	AR101984	AR101984	Sequence
25	1261.8	99.7	1343	6	BD064229	BD064229	Therapy f
26	1261.8	99.7	1343	6	BD247365	BD247365	Treatment
27	1261.8	99.7	1343	6	CS105027	CS105027	Sequence
28	1261.8	99.7	1343	6	AR234439	AR234439	Sequence
29	1261.8	99.7	1343	6	AR322462	AR322462	Sequence
30	1261.8	99.7	1343	6	AR372138	AR372138	Sequence
31	1261.8	99.7	1350	6	CO899396	CO899396	Sequence
32	1261.8	99.7	1350	6	CS033394	CS033394	Sequence
33	1261.8	99.7	1350	6	CS035758	CS035758	Sequence
34	1261.8	99.7	1350	6	CS042346	CS042346	Sequence
35	1261.8	99.7	1350	6	CS044710	CS044710	Sequence
36	1261.8	99.7	1390	6	AX780273	AX780273	Sequence
37	1261.8	99.7	1393	6	I30313	I30313	Sequence
38	1261.8	99.7	1393	6	AX030659	AX030659	Sequence
39	1261.8	99.7	1447	6	BD134457	BD134457	Human nuc
40	1261.8	99.7	1447	6	AX017286	AX017286	Sequence
41	1261.8	99.7	1447	6	AX524990	AX524990	Sequence
42	1260.2	99.5	1341	8	AK222627	AK222627	Homo sapi
43	1258.6	99.4	1350	6	CO882046	CO882046	Sequence
44	1257	99.3	1295	6	AX330543	AX330543	Sequence
45	1257	99.3	1295	8	HSAGALPS	HSAGALPS	Human alpha

ALIGNMENTS

RESULT 1
LOCUS AR653696 1266 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 11 from patent US 6887696.
ACCESSION AR653696
VERSION AR653696.1 GI:67584281
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 11 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
1..1266
/organism="unknown"
/mol_type="genomic DNA"

Query Match	Score	DB 6;	Length	1266;
Best Local Similarity	100.0%;	Pred. No. 0;	Matches 1266;	Conservative 0;
Matches 1266;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTTCCTGGCC	60	
Db	1	ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTTCCTGGCC	60	
Qy	61	CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGATTTGGCAAGGACGCT	120	
Db	61	CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGATTTGGCAAGGACGCT	120	
Qy	121	ACCATGGCTGGCTGGCGCTTTCATGTGCAACCTTGACTGCCAGAGAGGCCA	180	
Db	121	ACCATGGCTGGCTGGCGCTTTCATGTGCAACCTTGACTGCCAGAGAGGCCA	180	
Qy	181	GATTCTTCATCAGTGAAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTTCAAGGC	240	
Db	181	GATTCTTCATCAGTGAAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTTCAAGGC	240	

241 TGGAGGATGCGAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
Db |||||
241 TGGAGGATGCGAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
QY |||||
301 GATTGAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTTA 360
Db |||||
301 GATTGAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTTA 360
QY |||||
361 GCTAATTATGTTTACAGCAAGAGCTGAAGCTAGGATTTATGCGAGATGTTGGAATAAA 420
Db |||||
361 GCTAATTATGTTTACAGCAAGAGCTGAAGCTAGGATTTATGCGAGATGTTGGAATAAA 420
QY |||||
421 ACTGCGCAGGCTTCCTCGGAGTTTGGATACACTGCAATGATGCCAGACCTTTGCT 480
Db |||||
421 ACTGCGCAGGCTTCCTCGGAGTTTGGATACACTGCAATGATGCCAGACCTTTGCT 480
QY |||||
481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGGAAAAATTTG 540
Db |||||
481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGGAAAAATTTG 540
QY |||||
541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCGAGAGCAATGTGTAC 600
Db |||||
541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCGAGAGCAATGTGTAC 600
QY |||||
601 TCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
Db |||||
601 TCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
QY |||||
661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATTTGATGATGATTCCTGGAAAAATATAAG 720
Db |||||
661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATTTGATGATGATTCCTGGAAAAATATAAG 720
QY |||||
721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACAGGG 780
Db |||||
721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACAGGG 780
QY |||||
781 GGTGGAAATGA CCGAGATATGTAGTATGTCGCACTTTGGCTCAGCTGGAATCAGCAA 840
Db |||||
781 GGTGGAAATGA CCGAGATATGTAGTATGTCGCACTTTGGCTCAGCTGGAATCAGCAA 840
QY |||||
841 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
Db |||||
901 GCACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAAATGCCATCAAT 960
Db |||||
901 GCACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAAATGCCATCAAT 960
QY |||||
961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Db |||||
961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
QY |||||
1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAACCGGAGGAGATGGT 1080
Db |||||
1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAACCGGAGGAGATGGT 1080
QY |||||
1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
Db |||||
1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
QY |||||
1141 GCTGCTTCACTACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db |||||
1141 GCTGCTTCACTACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY |||||
1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTAGAAAAACACA 1260
Db |||||
1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTAGAAAAACACA 1260
QY |||||
1261 ATGTAA 1266
Db |||||
1261 ATGTAA 1266

RESULT 2
AR654964
LOCUS AR654964 1266 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 11 from patent US 6890748.
ACCESSION AR654964
VERSION AR654964.1 GI:67586503
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 11 10-MAY-2005;
FEATURES large Scale Biology Corporation; Vacaville, CA
location/Qualifiers
1..1266
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 100.0%; Score 1266; DB 6; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCGTTGCGCTTCCTGGCC 60
Db 1 ATGCAAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCGTTGCGCTTCCTGGCC 60
QY 61 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGACGCT 120
Db 61 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGACGCT 120
QY 121 ACCATGGGCTGCTGCACTGGGAGCGCTTCACTGTGCAACCTTTGACTGCCAGGAAAGCCA 180
Db 121 ACCATGGGCTGCTGCACTGGGAGCGCTTCACTGTGCAACCTTTGACTGCCAGGAAAGCCA 180
QY 181 GATTCCTCCATCAGTGAGAACTCTTCATGAGAGTGGCAGAGCTCATGGTCTCAGAGGC 240
Db 181 GATTCCTCCATCAGTGAGAACTCTTCATGAGAGTGGCAGAGCTCATGGTCTCAGAGGC 240
QY 241 TGGAGGATGCAAGTTATGATGATCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
Db 241 TGGAGGATGCAAGTTATGATGATCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
QY 301 GATTGAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTTA 360
Db 301 GATTGAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTTA 360
QY 361 GCTAATTATGTTTACAGCAAGAGCTGAAGCTAGGATTTATGCGAGATGTTGGAATAAA 420
Db 361 GCTAATTATGTTTACAGCAAGAGCTGAAGCTAGGATTTATGCGAGATGTTGGAATAAA 420
QY 421 ACCTCGCAGGCTTCCTCGGAGTTTGGATACACTGCAATGATGCCAGACCTTTGCT 480
Db 421 ACCTCGCAGGCTTCCTCGGAGTTTGGATACACTGCAATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGGAAAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGGAAAAATTTG 540
QY 541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCGAGAGCAATGTGTAC 600
Db 541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCGAGAGCAATGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATTTGATGATGATTCCTGGAAAAATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATTTGATGATGATTCCTGGAAAAATATAAG 720


```

QY 721 AGTATCTGGAGTGCACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
DB 721 AGTATCTTTGGAGTGCACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGSCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGSCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
DB 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGATAGGAGCTAAATTCGCATCAAT 960
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGATAGGAGCTAAATTCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGGAGACAACTTTGMAAGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGGAGACAACTTTGMAAGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGGACAGGATTTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGGACAGGATTTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATGCACTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATGCACTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAGCTAGGTTCTATGATGAGCT 1200
DB 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAGCTAGGTTCTATGATGAGCT 1200
QY 1201 TCAAGTTTAAAGTACATAAATCCACAGGCACTGTTTTCCTCAGCTAGAAAACACA 1260
DB 1201 TCAAGTTTAAAGTACATAAATCCACAGGCACTGTTTTCCTCAGCTAGAAAACACA 1260
QY 1261 ATGTAA 1266
DB 1261 ATGTAA 1266

RESULT 3
AR653697 LOCUS AR653697 1284 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 13 from patent US 6887696.
ACCESSION AR653697
VERSION AR653697.1 GI:67584282
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1284)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 13 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source 1..1284
/mol_type="genomic DNA"

ORIGIN
Query Match 99.8%; Score 1264; DB 6; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTGGCCTTTCGCTTCTCGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTGGCCTTTCGCTTCTCGGCC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGGATTTGGCAAGGAGCGCT 120
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGGATTTGGCAAGGAGCGCT 120

```

```

QY 121 ACCATGGGCTGGCTGCACACTGGAGCGCTTCATGTGCAACCTTGACTGCGAGAAAGCCA 180
DB 121 ACCATGGGCTGGCTGCACACTGGAGCGCTTCATGTGCAACCTTGACTGCGAGAAAGCCA 180
QY 181 GATTCTCTGCATCAGTGAAGCTCTTCATGAGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCTCTGCATCAGTGAAGCTCTTCATGAGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGGAGAGATGAGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300
DB 241 TGGAGAGATGAGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAAGAGGACACTTCAGGCGAGACCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAAGAGGACACTTCAGGCGAGACCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTATGTTACACGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAAATAA 420
DB 361 GCTAATTATGTTACACGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAAATAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGGCTTCCCTGGGAGTTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATGCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTG 540
DB 481 GACTGGGAGTAGATGCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTG 540
QY 541 GCAGATGGTTATAGCACATGCTCTGGCCCTGATAGGACTGCGCAGAGCAATTTGTGTAC 600
DB 541 GCAGATGGTTATAGCACATGCTCTGGCCCTGATAGGACTGCGCAGAGCAATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATATATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATATATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTCTGCACTGATGATTTCTGGAAGATATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTCTGCACTGATGATTTCTGGAAGATATAAG 720
QY 721 AGTATCTTGGAGTGCACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
DB 721 AGTATCTTGGAGTGCACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGSCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGSCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
DB 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGATAGGAGCTAAATTCGCCATCAAT 960
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGATAGGAGCTAAATTCGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGGAGACAACTTTGMAAGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGGAGACAACTTTGMAAGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGGACAGGATTTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGGACAGGATTTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATGCACTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATGCACTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAGCTAGGTTCTATGATGAGCT 1200
DB 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAGCTAGGTTCTATGATGAGCT 1200

```

```
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCAGCTCAGCTAGAAAACACA 1260
|
|
|
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCAGCTCAGCTAGAAAACACA 1260
|
|
|
QY 1261 ATGT 1264
|
|
|
Db 1261 ATGT 1264

RESULT 4
AR654965
LOCUS AR654965 1284 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 13 from patent US 6890748.
ACCESSION AR654965
VERSION AR654965.1 GI:67586504
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1284)
Garger,S.J., Turpen,T.H. and Kumagai,M.H.
AUTHORS Production of lysosomal enzymes in plants by transient expression
TITLE Patent: US 6890748-A 13 10-MAY-2005;
JOURNAL Large Scale Biology Corporation; Vacaville, CA
FEATURES
Location/Qualifiers
source
1..1284
/mol_type="genomic DNA"
ORIGIN
Query Match 99.8%; Score 1264; DB 6; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTTCGCTTTCCTGGCC 60
|
|
|
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTTCGCTTTCCTGGCC 60
|
|
|
QY 61 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGAATGGATTTGGCAAGAGCGCT 120
|
|
|
Db 61 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGAATGGATTTGGCAAGAGCGCT 120
|
|
|
QY 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCACTTGTACTGCCAGGAAGCA 180
|
|
|
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCACTTGTACTGCCAGGAAGCA 180
|
|
|
QY 181 GATTCCTGCATCAGTGAGAAGCTCTTCAATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
|
|
|
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCAATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
|
|
|
QY 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
|
|
|
Db 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
|
|
|
QY 301 GATTCAGAAGGAGAGCTTCAGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
|
|
|
Db 301 GATTCAGAAGGAGAGCTTCAGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
|
|
|
QY 361 GCTAATTTATGTTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAAGATTTGGAATAAA 420
|
|
|
Db 361 GCTAATTTATGTTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAAGATTTGGAATAAA 420
|
|
|
QY 421 ACTTGGCAGGCTTCCTGGGAGTTTGGATACATGACATTTGATGCTCCAGACCTTTGCT 480
|
|
|
Db 421 ACTTGGCAGGCTTCCTGGGAGTTTGGATACATGACATTTGATGCTCCAGACCTTTGCT 480
|
|
|
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
|
|
|
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
|
|
|
QY 541 GCAGATGGTTATAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTGAC 600
|
|
|
Db 541 GCAGATGGTTATAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTGAC 600
|
|
|
```

```
QY 601 TCCTGTGAGTGGCCCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
|
|
|
Db 601 TCCTGTGAGTGGCCCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
|
|
|
QY 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTCGAAAGATATAAG 720
|
|
|
Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTCGAAAGATATAAG 720
|
|
|
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
|
|
|
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
|
|
|
QY 781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCACTTTGGCCTCAGCTGGAATCAGNA 840
|
|
|
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCACTTTGGCCTCAGCTGGAATCAGNA 840
|
|
|
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTTAATGACCTC 900
|
|
|
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTTAATGACCTC 900
|
|
|
QY 901 CGACATCATGAGCTTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAAATTTGCCATCAAT 960
|
|
|
Db 901 CGACATCATGAGCTTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAAATTTGCCATCAAT 960
|
|
|
QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTACAGAGGGGAGACAACTTTGAAAGTGG 1020
|
|
|
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTACAGAGGGGAGACAACTTTGAAAGTGG 1020
|
|
|
QY 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCCGACAGAGATTGCT 1080
|
|
|
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCCGACAGAGATTGCT 1080
|
|
|
QY 1081 GGAACCTCCTTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
|
|
|
Db 1081 GGAACCTCCTTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
|
|
|
QY 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTAAGAGGAGCTAGGTTCTATGAAATGACT 1200
|
|
|
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTAAGAGGAGCTAGGTTCTATGAAATGACT 1200
|
|
|
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCAGCTAGAAAACACA 1260
|
|
|
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCAGCTAGAAAACACA 1260
|
|
|
QY 1261 ATGT 1264
|
|
|
Db 1261 ATGT 1264

RESULT 5
AR653694
LOCUS AR653694 1278 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 7 from patent US 6887696.
ACCESSION AR653694
VERSION AR653694.1 GI:67584279
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1278)
Garger,S.J., Turpen,T.H. and Kumagai,M.H.
AUTHORS Production of lysosomal enzymes in plants by transient expression
TITLE Patent: US 6887696-A 7 03-MAY-2005;
JOURNAL Large Scale Biology Corporation; Vacaville, CA
FEATURES
Location/Qualifiers
source
1..1278
/mol_type="genomic DNA"
ORIGIN
Query Match 99.8%; Score 1263.4; DB 6; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
```


Db	421	ACCTGCGCAGGCTTCCTCGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT	480
QY	481	GACTGGGGAGTAGACTCTGCTAAAATTTGATGGTTTACTGTGACACAGTTTGGAAAATTG	540
Db	481	GACTGGGGAGTAGACTCTGCTAAAATTTGATGGTTTACTGTGACACAGTTTGGAAAATTG	540
QY	541	GCAGATGGTTATAAGACATGCTCTTGGCCCTGGAATPAGCACTGGCAGAAAGCAATTGTTAC	600
Db	541	GCAGATGGTTATAAGACATGCTCTTGGCCCTGGAATPAGCACTGGCAGAAAGCAATTGTTAC	600
QY	601	TCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA	660
Db	601	TCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA	660
QY	661	CAGTACTGCAATCACATGGCGAAATTTGCTGCATTTGATGATTTCTTGGAAAAGTATAAAG	720
Db	661	CAGTACTGCAATCACATGGCGAAATTTGCTGCATTTGATGATTTCTTGGAAAAGTATAAAG	720
QY	721	AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG	780
Db	721	AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG	780
QY	781	GGTTGGAATGACCCAGATATTTAGTATTTGGCAACTTTTGGCCCTCAGCTGGAAATCAGCAA	840
Db	781	GGTTGGAATGACCCAGATATTTAGTATTTGGCAACTTTTGGCCCTCAGCTGGAAATCAGCAA	840
QY	841	GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTTATCATGCTTAATGACCTC	900
Db	841	GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTTATCATGCTTAATGACCTC	900
QY	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAAATTCGCATCAAT	960
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAAATTCGCATCAAT	960
QY	961	CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG	1020
Db	961	CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG	1020
QY	1021	GAACGACCTCTCTCAGGCTTAGCGCTGAGCTGATGATTAACCGCAGGAGATTGGT	1080
Db	1021	GAACGACCTCTCTCAGGCTTAGCGCTGAGCTGATGATTAACCGCAGGAGATTGGT	1080
QY	1081	GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAATCCT	1140
Db	1081	GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAATCCT	1140
QY	1141	GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT	1200
Db	1141	GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT	1200
QY	1201	TCAAAGTTAAGAGTCACATAAATCCACAGGCACCTGTTTTCCTTCAGCTAGAAAACACA	1260
Db	1201	TCAAAGTTAAGAGTCACATAAATCCACAGGCACCTGTTTTCCTTCAGCTAGAAAACACA	1260
QY	1261	ATGTA 1265	
Db	1261	ATGCA 1265	
RESULT 7			
LOCUS	AR653692	1290 bp	DNA linear PAT 13-JUN-2005
DEFINITION	Sequence 3 from patent US 6887696.		
ACCESSION	AR653692		
VERSION	AR653692.1	GI:67584277	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1290)		
AUTHORS	Garger,S.J., Turpen,T.H. and Kumagai,M.H.		
TITLE	Production of lysosomal enzymes in plants by transient expression		
JOURNAL	Patent: US 6887696-A 3 03-MAY-2005;		

Large Scale Biology Corporation; Vacaville, CA			
Location/Qualifiers			
1..1290			
/organism="unknown"			
/mol_type="genomic DNA"			
ORIGIN			
Query Match 99.8%; Score 1263.4; DB 6; Length 1290;			
Best Local Similarity 99.9%; Pred. No. 0;			
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	ATGCAAGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTTGCCTTCCCTTCTTGGCC	60
Db	1	ATGCAAGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTTGCCTTCCCTTCTTGGCC	60
QY	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACATGGAATGGAAGGAGCGCT	120
Db	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACATGGAATGGAAGGAGCGCT	120
QY	121	ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	180
Db	121	ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	180
QY	181	GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240
Db	181	GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240
QY	241	TGGAAGGATGCAAGTTATGATGACCTCTGCAATTTGATGATGCTGGATGGCTCCCCAAGA	300
Db	241	TGGAAGGATGCAAGTTATGATGACCTCTGCAATTTGATGATGCTGGATGGCTCCCCAAGA	300
QY	301	GATTCAGAAGGAGAGACTTCAGGCAGACCTTCAGCGCTTTCTTCATGGGATTCGCCAGCTA	360
Db	301	GATTCAGAAGGAGAGACTTCAGGCAGACCTTCAGCGCTTTCTTCATGGGATTCGCCAGCTA	360
QY	361	GCTAATTATGTTCA CAGCAAGGACTGAAGCTAGGGAATTTATGAGATGTTTGGAAATAAA	420
Db	361	GCTAATTATGTTCA CAGCAAGGACTGAAGCTAGGGAATTTATGAGATGTTTGGAAATAAA	420
QY	421	ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGCACTAGCATTCGCCAGACCTTTGCT	480
Db	421	ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGCACTAGCATTCGCCAGACCTTTGCT	480
QY	481	GACTGGGAGTAGATCTGCTAAAATTTGATGTTTACTGTGACAGATTTGGAAAATTG	540
Db	481	GACTGGGAGTAGATCTGCTAAAATTTGATGTTTACTGTGACAGATTTGGAAAATTG	540
QY	541	GCAGATGGTTATAAGACATGCTCTTGGCCCTGGAATPAGCACTGGCAGAAAGCAATTGTTAC	600
Db	541	GCAGATGGTTATAAGACATGCTCTTGGCCCTGGAATPAGCACTGGCAGAAAGCAATTGTTAC	600
QY	601	TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA	660
Db	601	TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA	660
QY	661	CAGTACTGCAATCACATGGCGAAATTTGCTGCATTTGATGATTTCTTGGAAAAGTATAAAG	720
Db	661	CAGTACTGCAATCACATGGCGAAATTTGCTGCATTTGATGATTTCTTGGAAAAGTATAAAG	720
QY	721	AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG	780
Db	721	AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG	780
QY	781	GGTTGGAATGACCCAGATATTTAGTATTTGGCAACTTTTGGCCCTCAGCTGGAAATCAGCAA	840
Db	781	GGTTGGAATGACCCAGATATTTAGTATTTGGCAACTTTTGGCCCTCAGCTGGAAATCAGCAA	840
QY	841	GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC	900
Db	841	GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC	900
QY	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAAATTCGCATCAAT	960
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAAATTCGCATCAAT	960

Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCCTTACAGGATAAGGACGTAAATGCCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGAGTACCAAGCTTAGACAGGAGAGCAAACTTTGAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGAGTACCAAGCTTAGACAGGAGAGCAAACTTTGAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGAGCTATGATAAACCGGAGAGATTTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGAGCTATGATAAACCGGAGAGATTTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCCAGTCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCCAGTCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGAAAGTACAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGAAAGTACAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGACCTGTTTTCCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGACCTGTTTTCCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 8
LOCUS AR654960 1290 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 3 from patent US 6890748.
ACCESSION AR654960
VERSION AR654960.1 GI:67586497
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Gargier, S.J., Turpen, T.H. and Kumagai, M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 3 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
Location/Qualifiers
source 1..1290
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 99.8%; Score 1263.4; DB 6; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCGAGCTGAGAAACCCAGAACTACATCTGGGCTGGGCGCTTGGCGCTTGGCTTCTGGCC 60
Db 1 ATGCGAGCTGAGAAACCCAGAACTACATCTGGGCTGGGCGCTTGGCGCTTGGCTTCTGGCC 60
Qy 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCT 120
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGCTGACCTGGAGGCGTTTCATGTGCAACCTTGA CTGCCAGGAAGGCA 180
Db 121 ACCATGGGCTGCTGACCTGGAGGCGTTTCATGTGCAACCTTGA CTGCCAGGAAGGCA 180
Qy 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
Db 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
Qy 241 TGGAGGATGCAAGGTTATCAGTACCTCTCATTTGATGACTGTTGATGGCTCCCAAGA 300
Db 241 TGGAGGATGCAAGGTTATCAGTACCTCTCATTTGATGACTGTTGATGGCTCCCAAGA 300
Qy 301 GATTTCAGAGGAGGAGCTTCAGGAGAGCCCTCAGCGCTTTCTTCATGGGATTCGCCAGCTA 360

RESULT 9
LOCUS AR653695
DEFINITION Sequence 9 from patent US 6887696.
ACCESSION AR653695

Db 301 GATTTCAGAGGAGAGCTTCAGGCGAGACCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATATGTTTCACAGCAAAAGGAGTAAAGCTTAGGGAATTTATGACAGATGTTGGAAATAAA 420
Db 361 GCTAATATGTTTCACAGCAAAAGGAGTAAAGCTTAGGGAATTTATGACAGATGTTGGAAATAAA 420
Qy 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATATACGACATTTGATGAGCAGTTTGGAAAATTTG 480
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATATACGACATTTGATGAGCAGTTTGGAAAATTTG 480
Qy 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTGTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTGTACTGTGACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGCTTATAAGCACATGCTTGGCCCTGAATAGGAGCTGGCAGAGCAATTTGTGTAC 600
Db 541 GCAGATGCTTATAAGCACATGCTTGGCCCTGAATAGGAGCTGGCAGAGCAATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCCCTCTTTTATATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCCCTCTTTTATATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTTCTCTGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTTCTCTGAAAAGTATAAG 720
Qy 721 AGTATCTTGGAGCTGGACATCTTTTAAACAGAGAGAAATTTGTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGAGCTGGACATCTTTTAAACAGAGAGAAATTTGTTGATGTTGCTGGACAGGG 780
Qy 781 GGTGGAAATGACCCAGAGATATCTAGTATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGAGATATCTAGTATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTGAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATCTCTAATGACCTC 900
Db 841 GTAACCTGAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATCTCTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAAGGATAAGGAGCTTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAAGGATAAGGAGCTTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGAGTACAGCTTAGACAGGAGAGCAAACTTTGAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGAGTACAGCTTAGACAGGAGAGCAAACTTTGAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGAGCTATGATAAACCGGAGAGATTTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGAGCTATGATAAACCGGAGAGATTTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGAGTGTCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGAGTGTCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGAAAGTACAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGAAAGTACAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGACCTGTTTTCCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGACCTGTTTTCCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

```

VERSION AR653695.1 GI:67584280
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1296)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 9 03-MAY-2005,
Large Scale Biology Corporation; Vacaville, CA
FEATURES
    source
        Location/Qualifiers
            1..1296
                /organism="unknown"
                /mol_type="genomic DNA"
ORIGIN
Query Match 99.8%; Score 1263.4; DB 6; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGNAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTCGCTTCCCTGGCC 60
DB 1 ATGCAGCTGAGNAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTCGCTTCCCTGGCC 60
QY 61 CTCGTTTCCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGAGCCCT 120
DB 61 CTCGTTTCCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGAGCCCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACTTGAATGCGCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACTTGAATGCGCAGGAAGGCCA 180
QY 181 GATTCCTGCATCAGTCAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
DB 181 GATTCCTGCATCAGTCAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
QY 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGAGTGGCTCCCCAAGA 300
DB 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGAGTGGCTCCCCAAGA 300
QY 301 GATTCAGAGGCAAGATTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
DB 301 GATTCAGAGGCAAGATTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
QY 361 GCTAATATGCTCACAGCAAGAGCTGAGCTAGGATTTATGCAGATGTTGGAAATAA 420
DB 361 GCTAATATGCTCACAGCAAGAGCTGAGCTAGGATTTATGCAGATGTTGGAAATAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATAGTACGACATTCATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATAGTACGACATTCATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTTGTTACTGTACAGATTTGGAAAATTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTTGTTACTGTACAGATTTGGAAAATTG 540
QY 541 GCAGATGTTATAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
DB 541 GCAGATGTTATAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAG 720
QY 721 AGTATCTTGAATGGAATCTTTTACAGGAGAGAAATTTGATGATTTCTGTCAGCAGGG 780
DB 721 AGTATCTTGAATGGAATCTTTTACAGGAGAGAAATTTGATGATTTCTGTCAGCAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCA 840
```

```

DB 781 GGTGGAAATGACCCAGATATGTTAGTGATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGCTATCATGCTGCTCTCTTTTATTTCATGCTAATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGCTATCATGCTGCTCTCTTTTATTTCATGCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACCTAATTCGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACCTAATTCGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGAGGTTA CCAGCTTAGACAGGGAGACA CTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGAGGTTA CCAGCTTAGACAGGGAGACA CTTTGAAGTGTGG 1020
QY 1021 GAACACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080
DB 1021 GAACACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAGGAAAGCTAGGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAGGAAAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTAGAAAACACA 1260
DB 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTAGAAAACACA 1260
QY 1261 ATGTA 1265
DB 1261 ATGCA 1265
RESULT 10
LOCUS AR654963
DEFINITION Sequence 9 from patent US 6890748.
ACCESSION AR654963
VERSION AR654963.1 GI:67586502
KEYWORDS
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1296)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 9 10-MAY-2005,
Large Scale Biology Corporation; Vacaville, CA
FEATURES
    source
        Location/Qualifiers
            1..1296
                /organism="unknown"
                /mol_type="genomic DNA"
ORIGIN
Query Match 99.8%; Score 1263.4; DB 6; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGNAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTCGCTTCCCTGGCC 60
DB 1 ATGCAGCTGAGNAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTCGCTTCCCTGGCC 60
QY 61 CTCGTTTCCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGAGCCCT 120
DB 61 CTCGTTTCCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGAGCCCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACTTGAATGCGCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACTTGAATGCGCAGGAAGGCCA 180
QY 181 GATTCCTGCATCAGTCAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
DB 181 GATTCCTGCATCAGTCAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
QY 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGAGTGGCTCCCCAAGA 300
DB 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGAGTGGCTCCCCAAGA 300
QY 301 GATTCAGAGGCAAGATTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
DB 301 GATTCAGAGGCAAGATTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
QY 361 GCTAATATGCTCACAGCAAGAGCTGAGCTAGGATTTATGCAGATGTTGGAAATAA 420
DB 361 GCTAATATGCTCACAGCAAGAGCTGAGCTAGGATTTATGCAGATGTTGGAAATAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATAGTACGACATTCATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATAGTACGACATTCATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTTGTTACTGTACAGATTTGGAAAATTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTTGTTACTGTACAGATTTGGAAAATTG 540
QY 541 GCAGATGTTATAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
DB 541 GCAGATGTTATAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAG 720
QY 721 AGTATCTTGAATGGAATCTTTTACAGGAGAGAAATTTGATGATTTCTGTCAGCAGGG 780
DB 721 AGTATCTTGAATGGAATCTTTTACAGGAGAGAAATTTGATGATTTCTGTCAGCAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCA 840
```


QY 661 CAGTACTGCAATCACCTGGCGAAATTTTGTGTCATTGATGATTTCTTGGAAAGTATAAAG 720
DB |||||
QY 661 CAGTACTGCAATCACCTGGCGAAATTTTGTGTCATTGATGATTTCTTGGAAAGTATAAAG 720
DB |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
DB |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
DB |||||
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGGAATGGCACTTTGGCCCTCAGCTGGATACGCAA 840
DB |||||
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGGAATGGCACTTTGGCCCTCAGCTGGATACGCAA 840
DB |||||
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGTCTCTTTATTCAATGCTTAATGACCTC 900
DB |||||
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGTCTCTTTATTCAATGCTTAATGACCTC 900
DB |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAAATGCCATCAAT 960
DB |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAAATGCCATCAAT 960
DB |||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
DB |||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
DB |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGGCAGAGATTGGT 1080
DB |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGGCAGAGATTGGT 1080
DB |||||
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
DB |||||
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
DB |||||
QY 1141 GCCTGCTTATACACAGCTCTCCCTGTGTAAGGAAGCACTAGGCTTCTATGAATGGACT 1200
DB |||||
QY 1141 GCCTGCTTATACACAGCTCTCCCTGTGTAAGGAAGCACTAGGCTTCTATGAATGGACT 1200
DB |||||
QY 1201 TCAAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
DB |||||
QY 1201 TCAAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
DB |||||
QY 1261 ATGTA 1265
DB |||||
QY 1261 ATGCA 1265
DB |||||

RESULT 12
AR654961
LOCUS AR654961 1308 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 5 from patent US 6890748.
ACCESSION AR654961
VERSION AR654961.1 GI:67586498
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1308)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 5 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source 1. .1308
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 99.8%; Score 1263.4; DB 6; Length 1308;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCCCTGGCC 60
DB |||||
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCCCTGGCC 60
DB |||||

QY 611 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTCGCAAGGACGCT 120
DB |||||
QY 611 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTCGCAAGGACGCT 120
DB |||||
QY 121 ACCATGGGCTGGCTGCATCTGGGAGGCTTCATGTGCAACCTTTGACTGCGCAGGAAGACCA 180
DB |||||
QY 121 ACCATGGGCTGGCTGCATCTGGGAGGCTTCATGTGCAACCTTTGACTGCGCAGGAAGACCA 180
DB |||||
QY 181 GATTCTCTGATCAGTGAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB |||||
QY 181 GATTCTCTGATCAGTGAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB |||||
QY 241 TGGAGGATGCAAGGTTATGAGTACCTCTGCAATGATGATGCTGTGATGCTGCCCAAGA 300
DB |||||
QY 241 TGGAGGATGCAAGGTTATGAGTACCTCTGCAATGATGATGCTGTGATGCTGCCCAAGA 300
DB |||||
QY 301 GATTTCAGAAAGCAGACTTCAGGCAGACCTCAGCGCTTTCTTCATGGGATTCGCCAGCTA 360
DB |||||
QY 301 GATTTCAGAAAGCAGACTTCAGGCAGACCTCAGCGCTTTCTTCATGGGATTCGCCAGCTA 360
DB |||||
QY 361 GCTAATTTATGTTCAAGCAAGGACTGAAGCTAGGATTTATGCAAGATTTGGAAATAAA 420
DB |||||
QY 361 GCTAATTTATGTTCAAGCAAGGACTGAAGCTAGGATTTATGCAAGATTTGGAAATAAA 420
DB |||||
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACGACATTTGATGCTGACAGTTTGGAAATTTG 540
DB |||||
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACGACATTTGATGCTGACAGTTTGGAAATTTG 540
DB |||||
QY 481 GACTGGGAGTAGATCTGCTAAAATTTTGTATGTTGTTTACTGTGACAGTTTGGAAATTTG 540
DB |||||
QY 481 GACTGGGAGTAGATCTGCTAAAATTTTGTATGTTGTTTACTGTGACAGTTTGGAAATTTG 540
DB |||||
QY 541 GCAGATGTTTAAAGCAATGCTCTGGCCCTGAAATAGGACTGGCAGAGCAATGTTGTAC 600
DB |||||
QY 541 GCAGATGTTTAAAGCAATGCTCTGGCCCTGAAATAGGACTGGCAGAGCAATGTTGTAC 600
DB |||||
QY 601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB |||||
QY 601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB |||||
QY 661 CAGTACTGCAATCACCTGGCGAAATTTTGTGTCATTGATGATTTCTTGGAAAGTATAAAG 720
DB |||||
QY 661 CAGTACTGCAATCACCTGGCGAAATTTTGTGTCATTGATGATTTCTTGGAAAGTATAAAG 720
DB |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
DB |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
DB |||||
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGGAATGGCACTTTGGCCCTCAGCTGGATACGCAA 840
DB |||||
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGGAATGGCACTTTGGCCCTCAGCTGGATACGCAA 840
DB |||||
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGTCTCTTTATTCAATGCTTAATGACCTC 900
DB |||||
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGTCTCTTTATTCAATGCTTAATGACCTC 900
DB |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAAATGCCATCAAT 960
DB |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAAATGCCATCAAT 960
DB |||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
DB |||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
DB |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGGCAGAGATTGGT 1080
DB |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGGCAGAGATTGGT 1080
DB |||||
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
DB |||||
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
DB |||||

Qy	1141	GCCTGCTTCATCACACAGCTCCTCCCTGTGTGAAAGGAAGCTAGGGTCTTATGAATGGACT	1200
Db	1141	GCCTGCTTCATCACACAGCTCCTCCCTGTGTGAAAGGAAGCTAGGGTCTTATGAATGGACT	1200
Qy	1201	TCRAGGTTAAGAAGTTCACATAAATCCACACAGGCACCTGTTTTCCTTCAGCTAGAAAACACA	1260
Db	1201	TCRAGGTTAAGAAGTTCACATAAATCCACACAGGCACCTGTTTTCCTTCAGCTAGAAAACACA	1260
Qy	1261	ATGTA 1265	
Db	1261	ATGCA 1265	
RESULT 13			
CS135670			
LOCUS	CS135670	1290 bp	DNA linear PAT 09-AUG-2005
DEFINITION	Sequence 1 from Patent WO2005069192.		
ACCESSION	CS135670		
VERSION	CS135670.1	GI:72056330	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Garman, S.C.		
TITLE	Crystal structure of human a-galactosidase		
JOURNAL	Patent: WO 2005069192-A 1 28-JUL-2005; Transkaryotic Therapies, Inc. (US); Seiden, Richard F. (US); Garboczi, David N. (US); Treco, Douglas A. (US); NIH/NTAID (US); Garman, Scott C. (US); Borowski, Marianne (US); Kinoshita, Carol M. (US)		
FEATURES			
source	Location/Qualifiers		
	1..1290		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	99.7%	Score 1261.8;	DB 6; Length 1290;
Best Local Similarity	99.8%;	Pred. No. 0;	
Matches 1263;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
Qy	1	ATGCAGCTGAGGAACCCAGAACATACATCTGGGCTGCGCGCTTGCGCTTCGCTTCCTGGCC	60
Db	1	ATGCAGCTGAGGAACCCAGAACATACATCTGGGCTGCGCGCTTGCGCTTCGCTTCCTGGCC	60
Qy	61	CTCGTTTCCTGGGACATCCTCGGGCTAGAGCACTGGACAATGGAATGGCAAGGACGCCT	120
Db	61	CTCGTTTCCTGGGACATCCTCGGGCTAGAGCACTGGACAATGGAATGGCAAGGACGCCT	120
Qy	121	ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	180
Db	121	ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	180
Qy	181	GATTTCCTGCATCAGTGAAGAAGCTCTTCAATGAGATGCGACAGCTCATGTGCTCAGAAAGC	240
Db	181	GATTTCCTGCATCAGTGAAGAAGCTCTTCAATGAGATGCGACAGCTCATGTGCTCAGAAAGC	240
Qy	241	TGGAAGGATGCAGGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA	300
Db	241	TGGAAGGATGCAGGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA	300
Qy	301	GATTTCAGAGGACAGCTTCAGGACAGCCCTCAGGCTTTTCCTCATGGGATTCGCCAGCTA	360
Db	301	GATTTCAGAGGACAGCTTCAGGACAGCCCTCAGGCTTTTCCTCATGGGATTCGCCAGCTA	360
Qy	361	GCTAATTATGTTACACAAAGGACTCAAGCTAGGGATTTATGCAGATGTTGGAATAA	420
Db	361	GCTAATTATGTTACACAAAGGACTCAAGCTAGGGATTTATGCAGATGTTGGAATAA	420
Qy	421	ACCTGCGCAGGCTTCCTCGGGAGTTTGGATATCTACGACATGATGCCAGACCTTTGCT	480

AUTHORS Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1290)
AUTHORS Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the Sali and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after Sali site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: <http://bioinfo.clontech.com/orfclones>.

FEATURES

source
 1..1290
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GH00112X1.0"
 /clone_lib="BD Creator (TM) CDS Library derived from MGC collection"
 /lab_host="DH5alpha T1 resistant"
 /note="Vector: pDNR-Dual"
 1..1290
 /codon_start=1
 /product="galactosidase, alpha"
 /protein_id="AAP35510.1"
 /gb_xref="GI:30582567"
 /translation="WQLRNPFLHLCALALFLALVSWDIPGARALDNLARTPMGWLHWRPMLDCQEPSCISEKLFEMAEMLVSEGKADGAYEICIDCWNPORDSEGLRQDPFRPHGIRQLYVHSLKGLIYADVGNKTCAGFPGSFGYDIDAQTFADNGVLLKPDGCTCSLENLADYKXMSLNRGTSIVYSCWPLYMPPFKPNYTEIRYCNHNFADIDDSWKSISILDTWSFNOERIVDVAGPGWMDPMLVIGNFGLSWNOQVTOMALMAIWAAPLMSNDLRHISPOKALLQDKDVIAINDPLGKQCYQLRQGDNFEWERPLSLANVAMINRQETGGPRSYTIAVASLKGVCNCPACFIQTQLLPVKRLKGFVWTSRLSRSHNPITGVLLQENTQMQLKDL"

ORIGIN

Query Match 99.7%; Score 1261.8; DB 8; Length 1290;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1263; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCCTTGGCC 60
 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCCTTGGCC 60
 61 CTGCTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGAGTGGCAAGAGCCCT 120
 61 CTGCTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGAGTGGCAAGAGCCCT 120
 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGGCAGGAAGGCCA 180
 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGGCAGGAAGGCCA 180
 181 GATTCCTGCATCAGTGAAGCTTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
 181 GATTCCTGCATCAGTGAAGCTTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
 241 TGGAAAGATGCAGGTTATGATGACCTTCGATGTGATGCTTGGATGGCTCCCAAGA 300
 241 TGGAAAGATGCAGGTTATGATGACCTTCGATGTGATGCTTGGATGGCTCCCAAGA 300
 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCCTCATGGGATTCGCCAGCTA 360

Db 301 GATTGAGAAGGCAGACTTCAGGCAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 Qy 361 GCTAATATATGTTACAGCAAGAGGCTGAGTCTGAGGATTTATGAGATGTTGGAAATAA 420
 Db 361 GCTAATATATGTTACAGCAAGAGGCTGAGTCTGAGGATTTATGAGATGTTGGAAATAA 420
 Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTCTTGGATATACAGCATTTGATGCCAGACCTTTGCT 480
 Db 421 ACCTGCGCAGGCTTCCCTGGGAGTCTTGGATATACAGCATTTGATGCCAGACCTTTGCT 480
 Qy 481 GACTGGGAGTAGATCTGCTAATAATTTGATGGTGTGTTACTGTGACAGTCTGAAAAATTG 540
 Db 481 GACTGGGAGTAGATCTGCTAATAATTTGATGGTGTGTTACTGTGACAGTCTGAAAAATTG 540
 Qy 541 GCAGATGTTATAGCAGATGCTTGGCCCTGAATAGAGCTGGCAGAGCAATTTGTGTAC 600
 Db 541 GCAGATGTTATAGCAGATGCTTGGCCCTGAATAGAGCTGGCAGAGCAATTTGTGTAC 600
 Qy 601 TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 Db 601 TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 Qy 661 CAGTACTGCAATCACTGGCGAAATTTTCTGTCATGATGATTTCTGGAAGATTAAG 720
 Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGTCATGATGATTTCTGGAAGATTAAG 720
 Qy 721 AGTATCTGGAGTGAATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACCGG 780
 Db 721 AGTATCTGGAGTGAATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACCGG 780
 Qy 781 GGTGGAATGACCCAGATATGTTAGTGTGGCACTTTGGCCTCAGCTGGAATCAGCA 840
 Db 781 GGTGGAATGACCCAGATATGTTAGTGTGGCACTTTGGCCTCAGCTGGAATCAGCA 840
 Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTATTCATGCTTAATGACCTC 900
 Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTATTCATGCTTAATGACCTC 900
 Qy 901 CGACATCAGCCCTCAAGCCCAAGCTCTCTTCAGGATTAAGAGTAAATTCGCCATCAAT 960
 Db 901 CGACATCAGCCCTCAAGCCCAAGCTCTCTTCAGGATTAAGAGTAAATTCGCCATCAAT 960
 Qy 961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTTAGACAGGGAGACAACTTTGAACTGTGG 1020
 Db 961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTTAGACAGGGAGACAACTTTGAACTGTGG 1020
 Qy 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
 Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
 Qy 1081 GGACCTCGCTCTTATACATGCGAGTTCCTCCCTGGGTAAGAGGATGGCTTAATCT 1140
 Db 1081 GGACCTCGCTCTTATACATGCGAGTTCCTCCCTGGGTAAGAGGATGGCTTAATCT 1140
 Qy 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGAGAGCTAGGCTTCTATGAATGGACT 1200
 Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGAGAGCTAGGCTTCTATGAATGGACT 1200
 Qy 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTCTCAGCTAGAAAAACA 1260
 Db 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTCTCAGCTAGAAAAACA 1260
 Qy 1261 ATGTA 1265
 Db 1261 ATGCA 1265

RESULT 15
 AY889631
 LOCUS
 DEFINITION
 AY889631 1290 bp mRNA linear SYN 29-MAR-2005
 Synthetic construct Homo sapiens clone FLH025769.01X galactosidase
 alpha (GLA) mRNA, complete cds.

[illegible]

Db 1261 ATGCA 1265

Search completed: December 26, 2005, 07:09:28
Job time : 6371 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 01:02:18 ; Search time 771 Seconds
(without alignments)
10943.581 Million cell updates/sec

Title: US-10-602-219-11

Perfect score: 1266

Sequence: 1 atgacgtctgaggaaccacaga.....agctagaacacacaaatgtaa 1266

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_21.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001as.*
- 5: Geneseq2001bs.*
- 6: Geneseq2002as.*
- 7: Geneseq2002bs.*
- 8: Geneseq2003as.*
- 9: Geneseq2003bs.*
- 10: Geneseq2003cs.*
- 11: Geneseq2003ds.*
- 12: Geneseq2004as.*
- 13: Geneseq2004bs.*
- 14: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	100.0	1266	6	AAD45223
2	1266	100.0	1266	10	ADD84750 Human rGA
3	1266	100.0	1266	12	ADJ88276 Human WT
4	1266	100.0	1266	12	ADM48680 Human wil
5	1266	100.0	1266	13	ADU66915 Human alp
6	1266	100.0	1266	14	AEA27444 Human alp
7	1264	99.8	1284	6	AAD45224
8	1264	99.8	1284	10	ADD84752 Human rGA
9	1264	99.8	1284	12	ADJ88278 Human WT
10	1264	99.8	1284	12	ADM48682 Human wil
11	1264	99.8	1284	13	ADU66917 Human alp
12	1264	99.8	1284	14	AEA27446 Human alp
13	1263.4	99.8	1278	6	AAD45221
14	1263.4	99.8	1278	10	ADD84746 Human alp
15	1263.4	99.8	1278	12	ADJ88272 Human rGA
16	1263.4	99.8	1278	12	ADM48676 Human wil
17	1263.4	99.8	1278	13	ADU66911 Human alp
18	1263.4	99.8	1278	14	AEA27440 Human alp
19	1263.4	99.8	1290	6	AAD45219 Human WT

20	1263.4	99.8	1290	10	ADD84742	Add84742 Human alp
21	1263.4	99.8	1290	12	ADJ88268	Adj88268 Human WT
22	1263.4	99.8	1290	12	ADM48672	Adm48672 Human wil
23	1263.4	99.8	1290	13	ADU66907	Adu66907 Human alp
24	1263.4	99.8	1290	14	AEA27436	Aea27436 Human alp
25	1263.4	99.8	1296	6	AAD45222	Aad45222 Human rGA
26	1263.4	99.8	1296	10	ADD84748	Add84748 Human alp
27	1263.4	99.8	1296	12	ADJ88274	Adj88274 Human WT
28	1263.4	99.8	1296	12	ADM48678	Adm48678 Human wil
29	1263.4	99.8	1296	13	ADU66913	Adu66913 Human alp
30	1263.4	99.8	1296	14	AEA27442	Aea27442 Human alp
31	1263.4	99.8	1308	6	AAD45220	Aad45220 Human WT-
32	1263.4	99.8	1308	10	ADD84744	Add84744 Human alp
33	1263.4	99.8	1308	12	ADJ88270	Adj88270 Human WT
34	1263.4	99.8	1308	12	ADM48674	Adm48674 Human wil
35	1263.4	99.8	1308	13	ADU66909	Adu66909 Human alp
36	1263.4	99.8	1308	14	AEA27438	Aea27438 Human alp
37	1261.8	99.7	1290	12	ADN49739	Adn49739 Human alp
38	1261.8	99.7	1290	13	ADU74415	Adu74415 Human alp
39	1261.8	99.7	1290	14	AEA3187	Aeb43187 Human alp
40	1261.8	99.7	1306	8	ABZ79777	Abz79777 Human alp
41	1261.8	99.7	1319	10	ACF80581	Acf80581 Human alp
42	1261.8	99.7	1339	10	ADH54464	Adh54464 Human alp
43	1261.8	99.7	1343	2	AAV31801	Aav31801 Nucleotid
44	1261.8	99.7	1343	3	AAA75434	Aaa75434 Nucleotid
45	1261.8	99.7	1343	3	AAA70685	Aaa70685 Human alp

ALIGNMENTS

RESULT 1

AAD45223

ID AAD45223 standard; DNA; 1266 BP.

XX AAD45223;

XX 27-DEC-2002 (first entry)

XX Human rGAL-8 DNA.

XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-8; Gene; ds.
XX Homo sapiens.

XX OS

XX FH Key Location/Qualifiers

FT CDS 1..1266

FT /*tag= a

FT /product= "Human rGAL-8 protein"

XX US2002088024-A1.

XX PD 04-JUL-2002.

XX 13-NOV-2001; 2001US-00993059.

XX 26-JUL-2000; 2000US-00626127.

XX (GARG/) GARGER S J.

XX (TURP/) TURPEN T H.

XX (KUNA/) KUMAGAI M H.

XX Garger SJ, Turpen TH, Kumagai MH;

XX WPI; 2002-681656/73.

XX P-PSDB; AAE28210.

XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal

XX storage diseases.

XX Claim 1; Page 42-44; 89pp; English.

CC The invention relates to human alpha-galactosidase truncated at the
 CC carboxy terminus and the production of enzymatically active recombinant
 CC human and animal lysosomal enzymes. The invention is useful for producing
 CC lysosomal enzymes for treating lysosomal storage diseases, producing
 CC altered or mutated proteins, enzymatically active or otherwise, to serve
 CC as precursors or substrates for further in vivo or in vitro processing to
 CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-8 DNA
 XX
 SQ

Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Query Match 100.0%; Score 1266; DB 6; Length 1266;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGGCTTGGCGCTTGGCTTCTGGCC 60
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGGCTTGGCGCTTGGCTTCTGGCC 60
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 121 ACCATGGGTGGTGCATCTGGGAGCGCTTCATGTGCAACCTTGTACTGCCAGGAAGGCA 180
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 121 ACCATGGGTGGTGCATCTGGGAGCGCTTCATGTGCAACCTTGTACTGCCAGGAAGGCA 180
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 181 GATTCCTGTCATCAGTGAGAGCTCTTCATGGAGAGGCGAGAGCTCATGTCTCAGAGGC 240
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 181 GATTCCTGTCATCAGTGAGAGCTCTTCATGGAGAGGCGAGAGCTCATGTCTCAGAGGC 240
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 301 GATTCAGAGGAGAGCTTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 301 GATTCAGAGGAGAGCTTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 361 GCTAATTATGTTACAGCAAGAGCTGAAAGCTAGGAGTTATGCAATGTTGGAATATA 420
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 361 GCTAATTATGTTACAGCAAGAGCTGAAAGCTAGGAGTTATGCAATGTTGGAATATA 420
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGATGCCAGACCTTTGCT 480
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGATGCCAGACCTTTGCT 480
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTG 540
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTG 540
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 541 CGAGATGTTATAGCAGATGCTTGGCCCTGATAGGACTGGCAGAGGATTTGTTAC 600
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 541 CGAGATGTTATAGCAGATGCTTGGCCCTGATAGGACTGGCAGAGGATTTGTTAC 600
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 601 TCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 601 TCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGTCATTTGATGATTTCTGGAAAAGTATAAG 720
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGTCATTTGATGATTTCTGGAAAAGTATAAG 720
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 721 AGTATCTTGGCTGGACATCTTTTAC CAGGAGAGAAATTTGTCATGCTGTCGACGAGG 780
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 721 AGTATCTTGGCTGGACATCTTTTAC CAGGAGAGAAATTTGTCATGCTGTCGACGAGG 780
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 781 GGTGGAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 781 GGTGGAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTTCATGCTTAATGACCTC 900
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTTCATGCTTAATGACCTC 900
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 901 CGACACATCAGCCCTCAAGCCCAAGAGCTCTCTTCAGGATAAGGACGTAATGGCCATCAAT 960
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 901 CGACACATCAGCCCTCAAGCCCAAGAGCTCTCTTCAGGATAAGGACGTAATGGCCATCAAT 960
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 961 CAGGACCCCTTGGGCAAGGAGTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 961 CAGGACCCCTTGGGCAAGGAGTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1021 GAAACACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1021 GAAACACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1081 GGACCTCTCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAACTCT 1140
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1081 GGACCTCTCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAACTCT 1140
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGTAAGGAAAGCTAGGTTCTATGAATGGACT 1200
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGTAAGGAAAGCTAGGTTCTATGAATGGACT 1200
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1201 TCAGGTTAAAGAGTCACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1201 TCAGGTTAAAGAGTCACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1261 ATGTAA 1266
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 QY 1261 ATGTAA 1266
 DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 RESULT 2
 ADD84750
 ID ADD84750 standard; DNA; 1266 BP.
 XX
 AC ADD84750;
 XX
 DT 29-JAN-2004 (first entry)
 XX Human alpha-galactosidase rGAL-8 DNA.
 XX
 DE Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
 KW enzyme replacement therapy; lysosomal disease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 CDS 1..1266
 FT /*tag= a
 FT /product= "Human rGAL-8"
 XX
 US2003106095-A1.
 PN
 XX
 PD 05-JUN-2003.
 XX
 XX 20-MAR-2002; 2002US-00103327.
 PF
 XX
 XX 26-JUL-2000; 2000US-00626127.
 PR
 XX 13-NOV-2001; 2001US-00993059.
 PR
 XX (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 PA (KUNAI/) KUMAGAI M H.
 XX
 XX Garger SJ, Turpen TH, Kumagai MH;
 PI
 XX WPI; 2003-801257/75.
 DR P-PSDB; ADD84751.
 XX
 PT New polynucleotide for producing active recombinant human and animal

PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
PI Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
DR P-PSDB; ADJ88277.
XX
PT New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX Claim 1; SEQ ID NO 11; 71pp; English.
XX
CC The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC DNA.
XX
SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
Query Match 100.0%; Score 1266; DB 12; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGGATGGCAAGGACGCT 120
Db 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGGATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTCACCTGCCAGGAGGCCA 180
Db 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTCACCTGCCAGGAGGCCA 180
QY 181 GATTCTTCATCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 181 GATTCTTCATCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 241 TGAAGATCCAGGTATGAGTACCTCTGATTCATGATGATGGATGGCTCCCAAAGA 300
Db 241 TGAAGATCCAGGTATGAGTACCTCTGATTCATGATGATGGATGGCTCCCAAAGA 300
QY 301 GATTTCAGAGGACAGCTTCAGGACAGCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
Db 301 GATTTCAGAGGACAGCTTCAGGACAGCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
QY 361 GCTAAATATGTTTACAGCAAGGACCTGAAGCTAGGATTTATGAGATTTGGAAATAA 420
Db 361 GCTAAATATGTTTACAGCAAGGACCTGAAGCTAGGATTTATGAGATTTGGAAATAA 420
QY 421 ACTTGGCAGGCTCCCTGGGATTTGGATACATGACATTCATGCCAGACCTTTGCT 480
Db 421 ACTTGGCAGGCTCCCTGGGATTTGGATACATGACATTCATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTTGTCTGACAGATTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTTGTCTGACAGATTTGGAAATTTG 540
QY 541 GCAGATGGTTTATAGCAATGTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
Db 541 GCAGATGGTTTATAGCAATGTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600

QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTCCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTCCTGGACAGGG 780
QY 781 GGTTCGAATGACCCAGATATGTTAGTATTTGGCACTTTGGCCCTCAGCTGGAAATCAGCA 840
Db 781 GGTTCGAATGACCCAGATATGTTAGTATTTGGCACTTTGGCCCTCAGCTGGAAATCAGCA 840
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGTCTCTTTATTCATGTCTAATGACCTC 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGTCTCTTTATTCATGTCTAATGACCTC 900
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGAGCTAATTCGCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGAGCTAATTCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGAGCAACTTTTGAAGTGTG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGAGCAACTTTTGAAGTGTG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGACGAGATTTGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGACGAGATTTGT 1080
QY 1081 GGACCTGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTATTCCT 1140
Db 1081 GGACCTGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTATTCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGAGAGCTAGGTTCTTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGAGAGCTAGGTTCTTATGAATGGACT 1200
QY 1201 TCAAGTTAAGAGTACATAAATCCACAGGACCTGTTTTGCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGTTAAGAGTACATAAATCCACAGGACCTGTTTTGCTTCAGCTAGAAAACACA 1260
QY 1261 ATGTAA 1266
Db 1261 ATGTAA 1266
RESULT 4
ADM48680
ID ADM48680 standard; DNA; 1266 BP.
XX
AC ADM48680;
XX
XT 03-JUN-2004 (first entry)
XX
DE Human wild type rGAL-8 DNA.
XX
KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW Gene; db.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT 1..1266
FT /*tag= a
FT /product= "GAL-8 protein"
XX
PN US2004023281-A1.

```

XX PD 05-FEB-2004.
XX PF 23-JUN-2003; 2003US-00602220.
XX PR 26-FEB-1988; 88US-00160766.
XX PR 26-FEB-1988; 88US-00160771.
XX PR 17-FEB-1989; 89US-00310881.
XX PR 22-OCT-1990; 90US-00600244.
XX PR 31-JUL-1992; 92US-00923692.
XX PR 30-DEC-1992; 92US-00997733.
XX PR 29-DEC-1993; 93US-00176414.
XX PR 19-JAN-1994; 94US-00184237.
XX PR 14-OCT-1994; 94US-00324003.
XX PR 21-MAY-1999; 99US-00316572.
XX PR 26-JUL-2000; 2000US-00626127.
XX PR 13-NOV-2001; 2001US-00993059.
XX (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX PI Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX WPI; 2004-142650/14.
XX DR P-PSDB; ADM48681.
XX PT New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT disease.
XX PS Disclosure; SEQ ID NO 11; 72pp; English.
XX CC The present invention relates to novel galactosidase (Gal) proteins such
CC as rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
CC of the present invention are useful for producing recombinant lysosomal
CC enzymes for enzyme replacement therapy for treating human and animal
CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human wild type rGAL-8 DNA used in the exemplification of the invention.
XX SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Query Match 100.0%; Score 1266; DB 12; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTGCGTTCTCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTGCGTTCTCTGGCC 60

QY 61 CTCGTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT 120
DB 61 CTCGTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT 120

QY 121 ACCATGGGCTGGCTGACCTGGGAGGCGTTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGGCTGGCTGACCTGGGAGGCGTTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180

QY 181 GATTCCTGCATCAGTCAGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
DB 181 GATTCCTGCATCAGTCAGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240

QY 241 TGGAGGATGCAGGTTATGATACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
DB 241 TGGAGGATGCAGGTTATGATACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300

QY 301 GATTCAGAGGAGCACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAGAGGAGCACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

```

```

QY 361 GCTAATTATGTTTCACAGCAAAAGGACTGAAGCTAGGGATTTATGTCAGATGTTTGGAAATAAA 420
DB 361 GCTAATTATGTTTCACAGCAAAAGGACTGAAGCTAGGGATTTATGTCAGATGTTTGGAAATAAA 420

QY 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGGACATTTGATGTCAGATGTTTGGAAATTTG 480
DB 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGGACATTTGATGTCAGATGTTTGGAAATTTG 480

QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTCTTACTGTCGACAGATTTGGAAAATTTG 540
DB 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTCTTACTGTCGACAGATTTGGAAAATTTG 540

QY 541 GCAGATGTTTATAAGCAACATGTCCTTGGCCCTGTAATAGCACTGGCAGAACATTTGTGTAC 600
DB 541 GCAGATGTTTATAAGCAACATGTCCTTGGCCCTGTAATAGCACTGGCAGAACATTTGTGTAC 600

QY 601 TCCTGTGATGGCCCTCTTTATATGTCGTCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB 601 TCCTGTGATGGCCCTCTTTATATGTCGTCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTTCTCTGGAAGATATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTTCTCTGGAAGATATAAG 720

QY 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATTTGTGATGTTGCTGGAACAGGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATTTGTGATGTTGCTGGAACAGGG 780

QY 781 GGTGGATGACCCAGATATGTTAGTATTTGGCAACTTTGGCCCTCAGCTGGATCAGCA 840
DB 781 GGTGGATGACCCAGATATGTTAGTATTTGGCAACTTTGGCCCTCAGCTGGATCAGCA 840

QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTATTATTCATGCTTAATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTATTATTCATGCTTAATGACCTC 900

QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAAATGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAAATGCCATCAAT 960

QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020

QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAAACGGCAGGAGATTGTT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAAACGGCAGGAGATTGTT 1080

QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140

QY 1141 GCCTGCTTCAACACACAGCTCTCTCTGTGAAAAGAGAGCTAGGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCAACACACAGCTCTCTCTGTGAAAAGAGAGCTAGGGTTCTATGAATGGACT 1200

QY 1201 TCNAGGTTAAGNAGTCATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260
DB 1201 TCNAGGTTAAGNAGTCATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260

QY 1261 ATGTAA 1266
DB 1261 ATGTAA 1266

```

```

RESULT 5
ADU66915
ID ADU66915 standard; DNA; 1266 BP.
XX
AC ADU66915;
XX
DT 10-FEB-2005 (first entry)

```


Db 1141 GCTGCTTATACACAGCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTCAGCTAGAAAACACA 1260
QY 1261 ATGT 1264
Db 1261 ATGT 1264

RESULT 9

ADJ88278
ID ADJ88278 standard; DNA; 1284 BP.
XX
AC
XX ADJ88278;
DT 06-MAY-2004 (first entry)
XX Human WT rGAL-8 R (galactosidase) DNA.
DE
XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
XX Fabry's disease; Gaucher's disease; human; gene; ds.
KW
XX
OS Homo sapiens.

XX Key Location/Qualifiers
FH 1. .1284
FT CDS /*tag= a
FT /product= "Human WT rGAL-8R protein"
XX

US2004016021-A1.

XX
PD 22-JAN-2004.
XX
PP 23-JUN-2003; 2003US-00602219.
XX

PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX

PA (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX

PI Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
DR P-PSDB; ADJ88279.
XX

XX New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX

Claim 1; SEQ ID NO 13; 71pp; English.

XX The invention relates to nucleotide encoding galactosidase (GAL). The

CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC DNA.
XX

SQ Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;

Query Match 99.8%; Score 1264; DB 12; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGGCGCTTGCCTTCCTTCCCTGGCC 60

Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGGCGCTTGCCTTCCTTCCCTGGCC 60

QY 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCT 120

Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCT 120

QY 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180

Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180

QY 181 GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

Db 181 GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 241 TGAAGGATGCAAGTTATGAGTACCTTCATCTGATGACTGTGGAGTCTCCCAAGA 300

Db 241 TGAAGGATGCAAGTTATGAGTACCTTCATCTGATGACTGTGGAGTCTCCCAAGA 300

QY 301 GATTCAAGAGGAGAGCTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

Db 301 GATTCAAGAGGAGAGCTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 361 GCTAATTATGTTTCACAGCAAGAGAGCTGAAAGTAGGATTTATGACAGATGTGGAATAAA 420

Db 361 GCTAATTATGTTTCACAGCAAGAGAGCTGAAAGTAGGATTTATGACAGATGTGGAATAAA 420

QY 421 ACCTGGCAGAGCTTCCTCGGAGTTTGGATACAGACATGATGCCAGACCTTTGCT 480

Db 421 ACCTGGCAGAGCTTCCTCGGAGTTTGGATACAGACATGATGCCAGACCTTTGCT 480

QY 481 GACTGGGAGTAGACTCTGCTAAATTTGATGTTTGTACTGTGACAGATTGTGAAAAATTG 540

Db 481 GACTGGGAGTAGACTCTGCTAAATTTGATGTTTGTACTGTGACAGATTGTGAAAAATTG 540

QY 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600

Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600

QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660

Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCAGCTGGGAAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAAG 720

Db 661 CAGTACTGCAATCAGCTGGGAAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAAG 720

QY 721 AGTATCTTGGAGTCATCTTTTAAACAGAGAGAAATTTGATGATTTGTCGCAAGAGG 780

Db 721 AGTATCTTGGAGTCATCTTTTAAACAGAGAGAAATTTGATGATTTGTCGCAAGAGG 780

QY 781 GGTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840

Db 781 GGTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840

QY 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTTCTAATGACCTC 900

Db 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTTCTAATGACCTC 900

QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGATAAGACGTAATTCATCAAT 960

901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAAATGCGCATCAAT	960
961	CAGGACCCCTTGGGCAAGCAAGGCTACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020
961	CAGGACCCCTTGGGCAAGCAAGGCTACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020
1021	GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGACAGGAGATTGGT	1080
1021	GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGACAGGAGATTGGT	1080
1081	GGACCTCGCTCTTATACCATCGCAGTGTCTCCCTGGGTAAAGGAGTGGCGCTCTAATCCT	1140
1081	GGACCTCGCTCTTATACCATCGCAGTGTCTCCCTGGGTAAAGGAGTGGCGCTCTAATCCT	1140
1141	GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT	1200
1141	GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT	1200
1201	TCAAGGTTAAGAAGTCAATATAATCCCAAGGCACCTGTTTTGCTTCAGCTAGAAAAACACA	1260
1201	TCAAGGTTAAGAAGTCAATATAATCCCAAGGCACCTGTTTTGCTTCAGCTAGAAAAACACA	1260
1261	ATGT 1264	
1261	ATGT 1264	

RESULT 10	
ADM48682	
ID	ADM48682 standard; DNA; 1284 BP.
XX	
XX	ADM48682;
XX	
DT	03-JUN-2004 (first entry)
XX	
DE	Human wild type rGAL-8R DNA.
XX	
KW	Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW	lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW	Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW	gene; ds.

XX	US2004023281-A1.	
PN		
XX		
PD	05-FEB-2004.	
XX		
XX	23-JUN-2003; 2003US-00602220.	
PF	26-FEB-1988; 88US-00160766.	
XX	26-FEB-1988; 88US-00160771.	
PR	17-FEB-1989; 89US-00310881.	
PR	22-OCT-1990; 90US-00600244.	
PR	31-JUL-1992; 92US-00923692.	
PR	30-DEC-1992; 92US-00997733.	
PR	28-DEC-1993; 93US-00176414.	
PR	19-JAN-1994; 94US-00184237.	
PR	14-OCT-1994; 94US-00324003.	
PR	21-MAY-1999; 99US-00316572.	
PR	26-JUL-2000; 2000US-00626127.	
PR	13-NOV-2001; 2001US-00993059.	
XX		
PA	(TURP/) TURPEN T H.	
PA	(KUMA/) KUMAGAI M H.	
PA	(FOGU/) FOGUE G P.	
PA	(ERWI/) ERWIN R L.	
PA	(GRIL/) GRILL L K.	

XX	Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
PI	WPI; 2004-142650/14.
XX	P-PSDB; ADM48683.
DR	
DR	
XX	
XX	New alpha-galactosidase polypeptides, useful in producing recombinant
PT	lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT	as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT	disease.
PT	
XX	
PS	Disclosure; SEQ ID NO 13; 72pp; English.
XX	
CC	The present invention relates to novel galactosidase (Gal) proteins such
CC	as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
CC	of the present invention are useful for producing recombinant lysosomal
CC	enzymes for enzyme replacement therapy for treating human and animal
CC	lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC	disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC	human wild type rGAL-8R DNA. This sequence comprises a human rGAL-8 DNA
CC	and ER retention signal DNA. This sequence is used in the exemplification
CC	of the invention.
XX	
SQ	Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;
Query Match	
Best Local Similarity 99.8%; Score 1264; DB 12; Length 1284;	
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCCTTCCTTGGGCC 60
DB	1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCCTTCCTTGGGCC 60
QY	61 CTCGTTTCTCGGCACATCCTCGGGCTAGAGCACTGGACAATGGATTGCCAAGACGCT 120
DB	61 CTCGTTTCTCGGCACATCCTCGGGCTAGAGCACTGGACAATGGATTGCCAAGACGCT 120
QY	121 ACATGGGCTGGCTGCACCTGGAGGCTTCATGTGCAACCTTGACTGCGCAGGAAGGCCA 180
DB	121 ACATGGGCTGGCTGCACCTGGAGGCTTCATGTGCAACCTTGACTGCGCAGGAAGGCCA 180
QY	181 GATTCTGTGCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB	181 GATTCTGTGCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY	241 TGGAAAGATGCAGGTTATGAGTACTCTGTGCATGTGATGATGCTTGGATGGCTCCCAAGA 300
DB	241 TGGAAAGATGCAGGTTATGAGTACTCTGTGCATGTGATGATGCTTGGATGGCTCCCAAGA 300
QY	301 GATTGAGAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB	301 GATTGAGAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY	361 GCTAAATTATGTTACAGCAAAAGGACTGAAAGCTAGGGATTTTATGCAGATGTTGGAAATAA 420
DB	361 GCTAAATTATGTTACAGCAAAAGGACTGAAAGCTAGGGATTTTATGCAGATGTTGGAAATAA 420
QY	421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATPACTACGACATGTGATGCCAGACCTTTGCT 480
DB	421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATPACTACGACATGTGATGCCAGACCTTTGCT 480
QY	481 GACTGGGGAGTACATCTGCTAAAATTTGATGGTTTGTACTCTGACAGTTTCGAAAATTG 540
DB	481 GACTGGGGAGTACATCTGCTAAAATTTGATGGTTTGTACTCTGACAGTTTCGAAAATTG 540
QY	541 GCAGATGGTTTATAAGCAATGCTTCCTGGCCCTGTAATAGGACTGGCAGAAAGCATTTGTGTAC 600
DB	541 GCAGATGGTTTATAAGCAATGCTTCCTGGCCCTGTAATAGGACTGGCAGAAAGCATTTGTGTAC 600
QY	601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTCAAAGGCCCAATTATACAGAAATCCGA 660
DB	601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTCAAAGGCCCAATTATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCACTGGCGAATTTTGTGACATGATGATTTCTTGGGAAAGTATAAAG 720
Db |||||||
QY 661 CAGTACTGCAATCACTGGCGAATTTTGTGACATGATGATTTCTTGGGAAAGTATAAAG 720
Db |||||||
QY 721 AGTATCTTGGACTGGACATCTTTTAACACGAGGAGAAATTTGTGATGTTGCTGGACACGG 780
Db |||||||
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTTGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Db |||||||
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTTGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Db |||||||
QY 841 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTATTATGCTTAATGACCTC 900
Db |||||||
QY 841 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTATTATGCTTAATGACCTC 900
Db |||||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db |||||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db |||||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db |||||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db |||||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGCGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
Db |||||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGCGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
Db |||||||
QY 1081 GGAACCTGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db |||||||
QY 1081 GGAACCTGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db |||||||
QY 1141 GCCTGCTTATACACAGCTCTCCCTGTGAAAGAGTCTAGGTTCTATGATGAGCT 1200
Db |||||||
QY 1141 GCCTGCTTATACACAGCTCTCCCTGTGAAAGAGTCTAGGTTCTATGATGAGCT 1200
Db |||||||
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAAACACA 1260
Db |||||||
QY 1261 ATGT 1264
Db |||||
QY 1261 ATGT 1264
Db |||||

RESULT 11
ADU66917
ID ADU66917 standard; DNA; 1284 BP.
XX
AC ADU66917;
XX
10-FEB-2005 (first entry)
XX
XX Human alpha-galactosidase protein encoding DNA #6.
XX
XX Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human; gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 1..1284
FT /*tag= a
FT /product= "Human alpha-galactosidase protein"
XX
XX US2004234516-A1.
XX
XX 25-NOV-2004.
XX
XX 21-MAY-2004; 2004US-00851388.
XX

PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
PR 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2004-821274/81.
XX P-PSDB; ADU66918.
XX
XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.
XX
XX Disclosure; SEQ ID NO 13; 88pp; English.
XX
XX The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein encoding DNA.
XX
XX Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;
SQ

Query Match 99.8%; Score 1264; DB 13; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTCCGCTTCTCTGGCC 60
Db |||||||
QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTCCGCTTCTCTGGCC 60
Db |||||||
QY 61 CTGCTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAGGAGCGCT 120
Db |||||||
QY 61 CTGCTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAGGAGCGCT 120
Db |||||||
QY 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATCTGCAACCTTGACCTCCAGGAAGAGCCA 180
Db |||||||
QY 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATCTGCAACCTTGACCTCCAGGAAGAGCCA 180
Db |||||||
QY 181 GATTCTCTGCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||||
QY 181 GATTCTCTGCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||||
QY 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATTGGCTCCCCAAGA 300
Db |||||||
QY 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATTGGCTCCCCAAGA 300
Db |||||||
QY 301 GATTTCAGAGGAGAGCTTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db |||||||
QY 301 GATTTCAGAGGAGAGCTTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db |||||||
QY 361 GCTAATTATGTTTCACAGCAAGGAGCTGAAGCTAGGATTTATGCAGATGTTGAAATATAA 420
Db |||||||
QY 361 GCTAATTATGTTTCACAGCAAGGAGCTGAAGCTAGGATTTATGCAGATGTTGAAATATAA 420
Db |||||||
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATTCATGCCCCAGACCTTTGCT 480
Db |||||||
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATTCATGCCCCAGACCTTTGCT 480
Db |||||||
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTACAGATTTGGAAAATTTG 540
Db |||||||
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTACAGATTTGGAAAATTTG 540
Db |||||||
QY 541 GCAGATGGTTATTAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
Db |||||||

Db	541	GCAGATGGTTATAAGACCATGTCTCTGGCCCTGAAATAGGACTGGCAGAAGCATTTGTGTAC	600
Qy	601	TCCGTGTAGTGGCCCTCTTTTATATGTGGCCCTTTCAAAGGCCAATTAATACAGAAATCCGA	660
Db	601	TCCGTGTAGTGGCCCTCTTTTATATGTGGCCCTTTCAAAGGCCAATTAATACAGAAATCCGA	660
Qy	661	CAGTACTGCAATCACTGCGGAAATTTTGTCTGACATTGATGATTCCTGGAAGAGTATAAAG	720
Db	661	CAGTACTGCAATCACTGCGGAAATTTTGTCTGACATTGATGATTCCTGGAAGAGTATAAAG	720
Qy	721	AGTATCTTGACTGGACATCTTTTAAACAGAGAGAAATGTGTGATGTCTGACACAGG	780
Db	721	AGTATCTTGACTGGACATCTTTTAAACAGAGAGAAATGTGTGATGTCTGACACAGG	780
Qy	781	GGTTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAAATCAGCAA	840
Db	781	GGTTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAAATCAGCAA	840
Qy	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGSGTCTCTCTTTTATTCATGTCTAATGACCTC	900
Db	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGSGTCTCTCTTTTATTCATGTCTAATGACCTC	900
Qy	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTTGCCATCAAT	960
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTTGCCATCAAT	960
Qy	961	CAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020
Db	961	CAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020
Qy	1021	GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGACAGGAGATTGGT	1080
Db	1021	GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGACAGGAGATTGGT	1080
Qy	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTCTTAATCCT	1140
Db	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTCTTAATCCT	1140
Qy	1141	GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT	1200
Db	1141	GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT	1200
Qy	1201	TCAAGGTTAAGAAAGTCAATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAAACA	1260
Db	1201	TCAAGGTTAAGAAAGTCAATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAAACA	1260
Qy	1261	ATGT 1264	
Db	1261	ATGT 1264	

RESULT 12	
AEA27446	
ID	AEA27446 standard; DNA; 1284 BP.
XX	
XX	AEA27446;
XX	
XX	11-2005 (first entry)
DT	
XX	
DE	Human alpha-galactosidase DNA, rGAL-8R, SEQ ID NO: 13.
XX	
KW	Gauchers disease; metabolic; neurological disease; niemann pick disease;
KW	genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
KW	antilepemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
KW	gene; ds.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
CDS	1..1284
FT	/*tag= a
FT	/product= "Human alpha-galactosidase protein"
FT	

XX	US2005125859-A1.
XX	
XX	09-JUN-2005.
XX	
XX	08-NOV-2004; 2004US-00984389.
XX	
XX	26-JUL-2000; 2000US-00626127.
XX	13-NOV-2001; 2001US-00993059.
XX	20-MAR-2002; 2002US-00103327.
XX	
XX	(LARG-) LARGE SCALE BIOLOGY CORP.
XX	
XX	Garger SJ, Turpen TH, Kumagai MH;
XX	
XX	WPI; 2005-404004/41.
XX	P-PSDB; AEA27447.
XX	
XX	New isolated polypeptides useful for producing lysosomal enzymes in
XX	plants to be utilized in enzyme replacement therapy or for the
XX	therapeutic treatment of human or animal lysosomal storage diseases, e.g.
XX	Gaucher's disease.
XX	
XX	Disclosure; SEQ ID NO 13; 88pp; English.
XX	
XX	The present invention relates to the production of human and animal
XX	lysosomal enzymes in plants by a transient plant expression system. The
XX	invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX	galactosidase (Gal) enzymes having a post-translational modification
XX	provided by the plant expression system. The invention is useful in
XX	enzyme replacement therapy for treating lysosomal storage diseases such
XX	as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX	disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX	also useful in researches for developing new approaches to medical
XX	treatment of lysosomal storage diseases and in industrial processes
XX	involving enzymatic substrate hydrolysis. The present sequence is the
XX	human alpha-galactosidase DNA.
XX	
XX	Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;

2

Db 361 GCTAATATATGTTACAGCAAGGACTGAAGCTAGGATTTATGACGATGTTGGAAATAAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATATACGACATTTGATGCCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATATACGACATTTGATGCCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGTTTATAAGCACATGTCCTTGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Db 541 GCAGATGTTTATAAGCACATGTCCTTGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCCCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCCCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGTCGACATTTGATGTCCTGGAAGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGTCGACATTTGATGTCCTGGAAGATATAAG 720
Qy 721 AGTATCTTTGGACTGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACCGGG 780
Db 721 AGTATCTTTGGACTGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACCGGG 780
Qy 781 GTTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GTTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACATGAGTGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACATGAGTGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAGGAGTAAATGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAGGAGTAAATGGCCATCAAT 960
Qy 961 CAGACCCCTTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAAGTGTGG 1020
Db 961 CAGACCCCTTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAAGTGTGG 1020
Qy 1021 GAAAGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080
Db 1021 GAAAGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080
Qy 1081 GGACCTGCTCTTATACCATCGCAGTCTCTCCCTGGCTTAAGGAGTGGCTGTAAATCCT 1140
Db 1081 GGACCTGCTCTTATACCATCGCAGTCTCTCCCTGGCTTAAGGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTTCATCACAGCTCTCTCCCTGTGAAAAGGAAAGTAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTTCATCACAGCTCTCTCCCTGTGAAAAGGAAAGTAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGTTAAGAAGTCAATAAATCCCAAGGCACTGTTTGTCTTCACTAGAAAACACA 1260
Db 1201 TCAAGTTAAGAAGTCAATAAATCCCAAGGCACTGTTTGTCTTCACTAGAAAACACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 14
ADD84746
ID ADD84746 standard; DNA; 1278 BP.
XX
AC ADD84746;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human alpha-galactosidase rGAL-4 DNA.
XX

KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..1278
FT /*tag= a
FT /product= "Human rGAL-4"
XX US2003106095-A1.
XX 05-JUN-2003.
XX 20-MAR-2002; 2002US-00103327.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2003-801257/75.
XX P-PSDB; ADD84747.
XX New polynucleotide for producing active recombinant human and animal
XX lysosomal enzymes in a plant expression system that can be used in enzyme
XX replacement therapy.
XX Claim 1; SEQ ID NO 7; 77pp; English.
XX The invention relates to human alpha-galactosidase derivatives and the
XX nucleic acids encoding them. The polypeptides are used in a method for
XX producing active recombinant human and animal lysosomal enzymes in a
XX plant expression system. The enzymes can be used in enzyme replacement
XX therapy for the therapeutic treatment of human and animal lysosomal
XX diseases. This sequence represents DNA encoding a human alpha-
XX galactosidase derivative polypeptide of the invention.
XX Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Query Match 99.8%; Score 1263.4; DB 10; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
Qy 61 CTGCTTTCTGGGACATCCCTGGGCTAGACACTGGACAATGGATTGCGAAGGACGCT 120
Db 61 CTGCTTTCTGGGACATCCCTGGGCTAGACACTGGACAATGGATTGCGAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTTCATGTGCAACCTTTGACTGCCAGGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTTCATGTGCAACCTTTGACTGCCAGGAGCCA 180
Qy 181 GATTCCTGTCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCCTGTCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTTGATGACTGTTGGATGGCTCCCCAAGA 300
Db 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTTGATGACTGTTGGATGGCTCCCCAAGA 300
Qy 301 GATTTCAGAGGACAGCTTCAGGCGAGACCTTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAGGACAGCTTCAGGCGAGACCTTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATATATGTTTCACAGCAAGGAGTGAAGCTAGGGAATTTATGACAGATGTTGGAAATAA 420

Db 361 GCTAATATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGTCAGATGTTGGAATAAA 420
Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGGACATTCAGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGGACATTCAGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTG 540
Qy 541 CGAGATGTTTATPAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Db 541 CGAGATGTTTATPAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Qy 601 TCTGTGAGTGGCTCTTTATATGATGCGCCCTTCAAAGCCCAATATACAGAAATCCGA 660
Db 601 TCTGTGAGTGGCTCTTTATATGATGCGCCCTTCAAAGCCCAATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTTGTGATGATTCCTGGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGATGATTCCTGGAAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Qy 781 GGTGGATGACCCAGATATGTTAGTATGTCGACATTTGGCCTCAGCTGGAAATCAGAA 840
Db 781 GGTGGATGACCCAGATATGTTAGTATGTCGACATTTGGCCTCAGCTGGAAATCAGAA 840
Qy 841 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCSCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCSCATCAAT 960
Qy 961 CAGGACCCCTTGGCAGCAAGGTTACAGCTTAGACGGGAGACAACTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGCAGCAAGGTTACAGCTTAGACGGGAGACAACTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCAGCAGAGATTGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCAGCAGAGATTGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
Qy 1141 GCTGCTTCATCACAGCTCTCCCTGTGAAAAGGAAGCTAGGTTCTATGAATGGACT 1200
Db 1141 GCTGCTTCATCACAGCTCTCCCTGTGAAAAGGAAGCTAGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGTTAAGAGTACATAAATCCACAGCAGCTGTTTTCCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGTTAAGAGTACATAAATCCACAGCAGCTGTTTTCCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 15
ID ADJ88272
XX ADJ88272 standard; DNA; 1278 BP.
XX
AC ADJ88272;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human rGAL- 4 (galactosidase) DNA.

XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease; human; gene; ds.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
CDS 1..1278
FT /*tag= a
FT /product= "Human WT rGAL-4 protein"
FT /partial
FT /note= "No start codon"
XX
PN US2004016021-A1.
XX
PD 22-JAN-2004.
XX
PF 23-JUN-2003; 2003US-00602219.
XX
PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX (TURP/) TURPEN T H.
PA (POGU/) POGUE G F.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
PI Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
DR P-PSDB; ADU88273.
XX
PT New lysosomal enzymes, useful in treating human and animal lysosomal storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
PS Claim 1; SEQ ID NO 7; 71pp; English.
XX
CC The invention relates to nucleotide encoding galactosidase (GAL). The invention is useful in gene therapy. The polynucleotides and polypeptides are useful in treating human and animal lysosomal storage diseases, e.g. Fabry's disease and Gaucher's diseases. The present sequence is human GAL DNA.
XX
SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;
Query Match 99.8%; Score 1263.4; DB 12; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCACTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTGCCTTCCTGGCC 60
Db 1 ATGCACTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTGCCTTCCTGGCC 60
Qy 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGACATGGATTGGCAAGGACGCT 120
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGACATGGATTGGCAAGGACGCT 120

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 04:17:59 ; Search time 4817 Seconds
(without alignments)
12296.543 Million cell updates/sec

Title: US-10-602-219-11
Perfect score: 1266
Sequence: 1 atgcagctgaggaaccacaga.....agctagaaacacaaatgtaa 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues
Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_est3.*
- 4: gb_hc.*
- 5: gb_est4.*
- 6: gb_est5.*
- 7: gb_est6.*
- 8: gb_est7.*
- 9: gb_gss1.*
- 10: gb_gss2.*
- 11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1261.8	99.7	1277	4	CR605654
2	1258.4	99.4	1266	4	CR607242
3	1248	98.6	1253	4	CR617861
4	1169.8	92.4	1290	10	AY408540
5	1149.4	90.8	1290	10	AY408541
6	968.6	76.5	1086	1	AL554978
7	926.8	73.2	1014	5	BX354096
8	922.8	72.9	999	1	AL552630
9	915.8	72.3	1067	3	BM564282
10	915.8	72.3	1026	1	AL577581
11	899.8	71.1	1376	4	AK054547
12	899.8	71.1	2962	4	AK040081
13	890	70.3	1133	7	CO645672
14	883.2	69.8	922	7	CO645623
15	880.8	69.6	1071	1	AL575861
16	874.6	69.1	1005	3	BQ662192
17	870	68.7	984	7	CO646251
18	862.4	68.1	920	7	CO645464
19	856.4	67.6	923	7	CT005156
20	828.6	65.5	1296	10	AY408542
21	814.2	64.3	976	5	BQ956043
22	807.8	63.8	937	5	BQ934640

23	797	63.0	922	5	BU540848	AGENCOURT
24	794.8	62.8	898	5	BU191867	AGENCOURT
25	783.8	61.9	888	5	BU154569	AGENCOURT
26	782	61.8	852	6	CA454083	AGENCOURT
27	768	60.7	801	5	BU596617	AGENCOURT
28	765.4	60.5	927	2	BE522583	601440703
29	765.2	60.4	813	5	BE522583	601440703
30	761.4	60.1	871	6	CA454143	AGENCOURT
31	757.6	59.8	1025	3	BM450649	AGENCOURT
32	752.2	59.4	958	6	CA487415	AGENCOURT
33	747.6	59.1	849	6	CA487531	AGENCOURT
34	747	59.0	796	8	DR156332	HESC2 64
35	747	59.0	845	7	CR983967	CR983967
36	746.8	59.0	908	2	BE379425	601237275
37	742.6	58.7	782	2	BI224248	602940538
38	742	58.6	871	3	BQ225444	AGENCOURT
39	741.4	58.6	1066	2	BG824323	602727346
40	741.2	58.5	921	6	CA488935	AGENCOURT
41	737.2	58.2	879	2	BG824387	602728627
42	729	57.6	749	5	BX344842	602728627
43	727.6	57.5	857	5	BX374627	602728627
44	713.8	56.4	800	6	CB529027	UI-H-FT2-
45	698.8	55.2	767	3	BI753664	603023559

ALIGNMENTS

RESULT 1
CR605654
LOCUS full-length cDNA clone CS0DB008YB11 of Neuroblastoma Cot
DEFINITION 10-normalized of Homo sapiens (human).
ACCESSION CR605654.1 GI:50486461
VERSION HTCC; CNSLT_CDNA.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1277)
AUTHORS Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Paradise Avenue
2 (bases 1 to 1277)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequençage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source location/Qualifiers
1..1277
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB008YB11"
/tissue_type="Neuroblastoma Cot 10-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 99.7%; Score 1261.8; DB 4; Length 1277;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1263; Conservative 0; Mismatches 2; Indels 0; Gaps 0;


```
QY 181 GATTCTCTGATCAGTGGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 187 GATTCTCTGATCAGTGGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 246
QY 241 TGGAAAGGATGCAAGGTTATGAGTACCTCTGCAATGATGATGCTGGTGGCTCCCAAGA 300
Db 247 TGGAAAGGATGCAAGGTTATGAGTACCTCTGCAATGATGATGCTGGTGGCTCCCAAGA 306
QY 301 GATTCAAGGAGAGCTTTCAGGCAGACCTCAGCGCTTTCCTCATCGGATTCGCCAGCTA 360
Db 307 GATTCAAGGAGAGCTTTCAGGCAGACCTCAGCGCTTTCCTCATCGGATTCGCCAGCTA 366
QY 361 GCTAATATATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAATATAA 420
Db 367 GCTAATATATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAATATAA 426
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGCATTGATGCGCCAGACCTTGGT 480
Db 427 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGCATTGATGCGCCAGACCTTGGT 486
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTCGAAATTTG 540
Db 487 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTCGAAATTTG 546
QY 541 CGAGATGTTTAAAGCAGATGCTCTGGCCCTGAATAGAGCTGGCAGAGCAATGTTGTAC 600
Db 547 CGAGATGTTTAAAGCAGATGCTCTGGCCCTGAATAGAGCTGGCAGAGCAATGTTGTAC 606
QY 601 TCTGTGAGTGGCTCTTATATGATGCGCCCTTCAAAAGCCCAATATACAGAAATCGA 660
Db 607 TCTGTGAGTGGCTCTTATATGATGCGCCCTTCAAAAGCCCAATATACAGAAATCGA 666
QY 661 CAGTACTGCAATCACTGGGCAATTTTGTGCAATTTGATGATTCCTGGAAAGTATAAG 720
Db 667 CAGTACTGCAATCACTGGGCAATTTTGTGCAATTTGATGATTCCTGGAAAGTATAAG 726
QY 721 AGTATCTTGGATGAGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db 727 AGTATCTTGGATGAGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 786
QY 781 GGTGGATGACCCAGATATGTTAGTATGTCGAATTTGATGATTCCTGGAAAGTATAAG 840
Db 787 GGTGGATGACCCAGATATGTTAGTATGTCGAATTTGATGATTCCTGGAAAGTATAAG 846
QY 841 GTAATCTCAGATGGCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTAATGACCTC 900
Db 847 GTAATCTCAGATGGCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTAATGACCTC 906
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTAATTCGCAATCAAT 960
Db 907 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTAATTCGCAATCAAT 966
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTACAGGAGAGACAACTTTGAAGTGTGG 1020
Db 967 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTACAGGAGAGACAACTTTGAAGTGTGG 1026
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGATGCTATGATAAACCAGGAGAGATGGT 1080
Db 1027 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGATGCTATGATAAACCAGGAGAGATGGT 1086
QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Db 1087 GGACCTCGCTCTTATACCATCGCAGTTGCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1146
QY 1141 GCCTGTTTCATCAGAGCTCTCCCTGTGAAAGAGAGCTAGGTTTCTATGAATGGACT 1200
Db 1147 GCCTGTTTCATCAGAGCTCTCCCTGTGAAAGAGAGCTAGGTTTCTATGAATGGACT 1206
QY 1201 TCAAGGTTAAGAGTACATNAATCCACAGGACCTGTTTTCCTTCAGCTAGAAACACA 1260
Db 1207 TCAAGGTTAAGAGTACATNAATCCACAGGACCTGTTTTCCTTCAGCTAGAAATACA 1266
```

```
RESULT 3
LOCUS CR617861
DEFINITION full-length cDNA clone CS0DK009YE12 of HeLa cells Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR617861
VERSION CR617861.1 GI:50498668
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 1253)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
source
1. 1253
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK009YE12"
/tissue type="HeLa cells Cot 25-normalized"
/plasmid="pCMVSPORT_6"
```

ORIGIN

```
Query Match 98.6%; Score 1248; DB 4; Length 1253;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGCCTTCGCTTCCTGGCC 60
Db 6 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGCCTTCGCTTCCTGGCC 65
QY 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Db 66 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 125
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAATCCAGGAGAGCA 180
Db 126 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAATCCAGGAGAGCA 185
QY 181 GATTCCTGCATCAGTGAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 186 GATTCCTGCATCAGTGAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGC 245
QY 241 TGGAAAGGATGAGGTTATGAGTACCTCTGCAATTCATGATGCTGGTGGCTCCCAAGA 300
Db 246 TGGAAAGGATGAGGTTATGAGTACCTCTGCAATTCATGATGCTGGTGGCTCCCAAGA 305
QY 301 GATTCAAGGAGAGCTTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 306 GATTCAAGGAGAGCTTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 365
QY 361 GCTAATATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAATATAA 420
Db 366 GCTAATATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAATATAA 425
```


QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATTCCTGGAAAGTATAAG 720
QY 721 AGTATCTGGAGTGCACATCTTTTAAACAGGAGAGAAATTTGTGCTGGACACGGG 780
Db 721 AGTATCTGGAGTGCACATCTTTTAAACAGGAGAGAAATTTGTGCTGGACACGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTGTGATGAGTGGCACTTTGGCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTGTGATGAGTGGCACTTTGGCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTGAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTGAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACATCATGAGCCCTCAGCCAAAGCTCTCTTACAGATAGGAGTAAATGCCATCAAT 960
Db 901 CGACATCATGAGCCCTCAGCCAAAGCTCTCTTACAGATAGGAGTAAATGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTACAGAGGAGAGCAACTTTGAAAGTG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTACAGAGGAGAGCAACTTTGAAAGTG 1020
QY 1021 GAAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
Db 1021 GAAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACATGACAGTGGTTCCTGGGTAAGGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACATGACAGTGGTTCCTGGGTAAGGAGTGGCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAGAGCTAGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAGAGCTAGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAAACA 1260
Db 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAAACA 1260
QY 1261 ATGTA 1265
Db 1261 ATGTA 1265

RESULT 5
AY408541
LOCUS
DEFINITION
AY408541 1290 bp DNA linear GSS 15-DEC-2003
Pan troglodytes GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION
AY408541
VERSION
AY408541.1 GI:39764512
KEYWORDS
GSS.
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
REFERENCE
1 (bases 1 to 1290)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
JOURNAL
PUBMED
14671302
REFERENCE
2 (bases 1 to 1290)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.

TITLE
JOURNAL
COMMENT
FEATURES
source
gene
ORIGIN
Query Match 90.8%; Score 1149.4; DB 10; Length 1290;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 1153; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTCGTTCCTGGCC 60
Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTCGTTCCTGGCC 60
QY 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGAATGGCAAGGACGCT 120
Db 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGAATGGCAAGGACGCT 120
QY 121 ACCATGGGCTGGCTGACCTGGGAGGCTTCTGTCGCACTTGGCTGGCAAGGACGCA 180
Db 121 ACCATGGGCTGGCTGACCTGGGAGGCTTCTGTCGCACTTGGCTGGCAAGGACGCA 180
QY 181 GATTCCTGTCATCAGTGAAGACTCTTTCATGAGATGGCAGAGCTCATGGTCTCNGAAGGC 240
Db 181 GATTCCTGTCATCAGTGAAGACTCTTTCATGAGATGGCAGAGCTCATGGTCTCNGAAGGC 240
QY 241 TGGAAAGATGAGGTTATGAGTACCTCTGCAATGATGATCTGTTGGATGGCTCCCAAGA 300
Db 241 TGGAAAGATGAGGTTATGAGTACCTCTGCAATGATGATCTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAGAGGACAGACTTCAGGAGAGCTTCAGGAGCTTCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAGAGGACAGACTTCAGGAGAGCTTCAGGAGCTTCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTTATGTTTCAGCAAGAGAGCTGAAGCTAGGATTTATGCAAGATTTGGAAATAA 420
Db 361 NNNNNNNNGTTTCAGCAAGAGAGCTGAAGCTAGGATTTATGCAAGATTTGGAAATAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATATACAGACATGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATATACAGACATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTTGCATGTTGTTACTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTTGCATGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 541 GCAGATGTTTATAAGACACATGCTCCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTG 600
Db 541 GCAGATGTTTATAAGACACATGCTCCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTG 600
QY 601 TCCTGTGAGTGGCTCTTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 NNN 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATTCCTGGAAAGTATAAG 720
QY 721 AGTATCTTGGAGTGCACATCTTTTAAACAGGAGAGAAATTTGTGCTGGACACGGG 780
Db 721 AGTATCTTGGAGTGCACATCTTTTAAACAGGAGAGAAATTTGTGCTGGACACGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTGTGATGAGTGGCACTTTGGCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTGTGATGAGTGGCACTTTGGCTCAGCTGGAATCAGCAA 840

Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
Location/Qualifiers
1..1290
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1290
/gene="GLA"
/locus_tag="HCM3258"

Db	781	GGTTGGAAATGACCAGATATGTTAGTGAATGGCGAACTTTGGCCCTCAGCTGGAATCAGCAA	840
Qy	841	GTAACTCAGATGGCCCTCTGGGCTATCATGGCTCTCTCTTTATTTCATGTCCTAATGACCTC	900
Db	841	GTAACTCAGATGGCCCTCTGGGCTATCATGGCTCTCTCTTTATTTCATGTCCTAATGACCTC	900
Qy	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGTAAGGACGCTAAATGGCCATCAAT	960
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGTAAGGACGCTAAATGGCCATCAAT	960
Qy	961	CAGGACCCCTTTGGCGAAGCAAGGCTACAGCTTAGACAGGGAGACAACTTTTGAAGTGTGG	1020
Db	961	CAGGACCCCTTTGGCGAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTTGAAGTGTGG	1020
Qy	1021	GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGCGAGGAGATTGGT	1080
Db	1021	GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGCGAGGAGATTGGT	1080
Qy	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTTAAAGGAGTGGCTGTAAATCCT	1140
Db	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTTAAAGGAGTGGCTGTAAATCCT	1140
Qy	1141	GCCTGCTTCATCAACAGCTCTCCTCTGTGAAAAAGGAAGCTAGGGTTCTATGAATGGAAT	1200
Db	1141	GCCTGCTTCATCAACAGCTCTCCTCTGTGAAAAAGGAAGCTAGGGTTCTATGAATGGAAT	1200
Qy	1201	TCAAGGTTAAGAGTCAATAAATCCCAAGGCACTGTTTTGCTTCAGCTAGAAAAACACA	1260
Db	1201	TCAAGGTTAAGAGTCAATAAATCCCAAGGCACTGTTTTGCTTCAGCTAGAAAAATACA	1260
Qy	1261	ATGTA 1265	
Db	1261	ATGCA 1265	

RESULT 6	AL554978	1086 bp	linear	EST 30-MAR-2004
LOCUS	AL554978			
DEFINITION	AL554978 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens			
	CDNA clone CS0DR009YE12 5-PRIME, mRNA sequence.			
ACCESSION	AL554978			
VERSION	AL554978.3	GI:45859723		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
	Hominidae; Homo.			
REFERENCE	1 (bases 1 to 1086)			
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished (2001)			
COMMENT	On Feb 15, 2001 this sequence version replaced gi:31276787.			

FEATURES	source
<pre> "acp...":1086 Location/Qualifiers 1,1086 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DK009YE12" /cell_type="HELA CELLS COT 25-NORMALIZED" /cell_line="HELA" </pre>	<pre> 841 GGTTGGAAATGACCCAGCATATGTTAGTCATTTGGCAAACCTTTGGCCCTCAGCTGGAATCAGCAA 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 16</pre>

```

Db      905  CCGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGAGACTWATGCCATCAA  964
Qy      960  TCAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTTAGACAGGAGACAACTTTTGAAGTGTC  1019
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      965  TCAGGA-CCCVTKGCCAAGCAAGGTA-CAGCTTAGAWAGGGRGACAMTTTGWAGTGTG  1022
Qy      1020  GGAACGACCTCTC 1032
      |||:|||||
Db      1023  GGAMGACTCTCTC 1035

RESULT 7
BX354096 1014 bp mRNA linear EST 23-APR-2004
LOCUS BX354096 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DC014YH04 5-PRIME, mRNA sequence.
ACCESSION BX354096
VERSION BX354096.2 GI:46550107
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE 1. (bases 1 to 1014)
AUTHORS Li,W.B., Gruber,C., Jesses,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 5, 2003 this sequence version replaced gi:30379845.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdnas-CS0DC014DD02QP1&c=10506.r.

FEATURES
source
    1..1014
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CS0DC014YH04"
        /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
        /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
        /note="1st strand cDNA was primed with a NotI-oligo(dT)
        primer. Five prime end enriched, double-strand cDNA was
        digested with Not I and cloned into the Not I and EcoR V
        sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 73.2%; Score 926.8; DB 5; Length 1014;
Best Local Similarity 98.4%; Pred. No. 2 8e-266;
Matches 958; Conservative 8; Mismatches 5; Indels 3; Gaps 3;

Qy      11  GGAACCCAGAACTACATCTGGCTGCGCGTTCGCTTCCTGCGCCCTCGTTTCCT  70
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      29  GGAACCCAGAACTACATCTGGCTGCGCGTTCGCTTCCTGCGCCCTCGTTTCCT  87
Qy      71  GGGNATCCCTGGGCTAGACCTGACCAATGGATTGGCAAGCGCCTACCATGGCT  130
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      88  GGGACATCTCCCTGGGCTAGACCTGACCAATGGATTGGCAAGCGCCTACCATGGCT  147
Qy      131  GGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCAGATTCCTGCA  190
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      148  GGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCAGATTCCTGCA  207
Qy      191  TCAGTGAGAAGCTCTTCATGGAGATGCCAGAGCTCATGGTCTCAGAGGCGTGAAGGATG  250
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Contact: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 10506.r

For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?c=CS0DI067CE01QP1&c=10506.r.

FEATURES

Source

1..999
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DI067YJ01"
 /issue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo (dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 72.9%; Score 922.8; DB 1; Length 999;
 Best Local Similarity 99.3%; Pred. NO. 4.4e-265;
 Matches 976; Conservative 3; Mismatches 0; Indels 5; Gaps 5;

QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTCTCGCTTCTCTGGCC 60
 DB 7 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTCTCGCTTCTCTGGCC 65

QY 61 CTCGTTTCTCTGGGACATCCTCTGGGCTAGAGCACTGGACAATGATGGCAAGGACGCT 120
 DB 66 CTCGTTTCTCTGGGACATCCTCTGGGCTAGAGCACTGGACAATGATGGCAAGGACGCT 125

QY 121 ACCATGGGCTGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACCTGCCAGGAGGCCA 180
 DB 126 ACCATGGGCTGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACCTGCCAGGAGGCCA 185

QY 181 GATTCTCTGATCAGTACAGAACTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGC 240
 DB 186 GATTCTCTGATCAGTACAGAACTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGC 245

QY 241 TGGAGGATGCAAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 246 TGGAGGATGCAAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 305

QY 301 GATTTCAGAGGACAGCTTCAGGAGACCTTCAGGAGACCTTCAGGAGACCTTCAGGAGACCT 360
 DB 306 GATTTCAGAGGACAGCTTCAGGAGACCTTCAGGAGACCTTCAGGAGACCTTCAGGAGACCT 365

QY 361 GCTAATATGTTTCAAGCAAGGACCTGAAGCTAGGATTTATGACATGTTGGAAATAAA 420
 DB 366 GCTAATATGTTTCAAGCAAGGACCTGAAGCTAGGATTTATGACATGTTGGAAATAAA 425

QY 421 ACTGCGCAGGCTTCCCTGGGAGTTTGGATGATGATGATGATGATGATGATGATGATGATGAT 480
 DB 426 ACTGCGCAGGCTTCCCTGGGAGTTTGGATGATGATGATGATGATGATGATGATGATGATGAT 485

QY 481 GACTGGGGAGTAGATCTGCTAAATTTGATGTTGTTTCTGCTGACATGTTGGAAATTTG 540
 DB 486 GACTGGGGAGTAGATCTGCTAAATTTGATGTTGTTTCTGCTGACATGTTGGAAATTTG 545

QY 541 GCAGATGTTTATAGACACATGCTCTGGGCTGCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
 DB 546 GCAGATGTTTATAGACACATGCTCTGGGCTGCTGAATAGGACTGGCAGAGCATTTGTGTAC 605

QY 601 TCTGTGAGTGGCTCTTTATATGTTGGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 DB 606 TCTGTGAGTGGCTCTTTATATGTTGGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 665

QY 661 CAGTACTGCAATCACTGGCGAAATTTTCTGCTGACATTTGATGATTTCTTGAAAGTATAAG 720
 DB 666 CAGTACTGCAATCACTGGCGAAATTTTCTGCTGACATTTGATGATTTCTTGAAAGTATAAG 725

QY 721 AGTATCTTGGAGCTGGACATCTTTTAAACAGGAGAGAATTTGTGATGTTCTGGACGAGG 780
 DB 726 AGTATCTTGGAGCTGGACATCTTTTAAACAGGAGAGAATTTGTGATGTTCTGGACGAGG 785

QY 781 GGTGGATGACCCAGATATGTTAGTAGTATGGCACTTTGGCTCAGCTGGAATCAGCAA 840
 DB 786 GGTGGATGACCCAGATATGTTAGTAGTATGGCACTTTGGCTCAGCTGGAATCAGCAA 845

QY 841 GTAACCTCAGATGGCCCTCTGGCTATCATGCTGCTCTTTATTCATGCTTAATGACCTC 900
 DB 846 GTAACCTCAGATGGCCCTCTGGCTATCATGCTGCTCTTTATTCATGCTTAATGACCTC 904

QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTCCTCATCAAT 960
 DB 905 CGACACATCAGCCCTCAAGCC-AGCTCTCTCTTTCAGGATAAGGACGTAATTC-CAWCAAT 962

QY 961 CAGGACCCCTTGGGCAAGCAAGGG 984
 DB 963 CAGGA-CCCTTGGGCAAGCAAGGK 985

RESULT 9
 BM564282
 LOCUS
 DEFINITION
 5', mRNA sequence.
 BM564282.1 GI:18811955
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo;
 1 (bases 1 to 1067)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12759 row: p column: 07
 High quality sequence stop: 761.

FEATURES

source

1..1067
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5742198"
 /tissue_type="medulla"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_119"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: EcoRV (destroyed); RNA source normal medulla from
 anonymous male age 27. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.3 kb, insert size range
 0.9-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 013. Note:
 this is a NIH_MGC Library."

ORIGIN

Query Match	72.3%	Score 915.8	DB 3	Length 1067
Best Local Similarity	96.4%	Pred. No. 5.7e-263		
Matches 1009	Conservative 0	Mismatches 31	Indels 7	Gaps 7
QY	6	GCTGAGAAACCCAGAACTACATCTGGGCTGCGCTTGGCGCTTCGCTTCGCTTCCTGGCCCTCGT	65	
DB	15	GCTGAGAAACCCAGAACTACATCTGGGCTGCGCGCTTGGCGCTTCGCTTCCTGGCCCTCGT	74	
QY	66	TTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCCCTACCAT	125	
DB	75	TTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCCCTACCAT	134	
QY	126	GGGCTGGCTGGACCTGGGAGCGCTTCATGTGCAACTCTTGCACTGCCAGGAAGCCAGATTC	185	
DB	135	GGGCTGGCTGGACCTGGGAGCGCTTCATGTGCAACTCTTGCACTGCCAGGAAGCCAGATTC	194	
QY	186	CTGCATCAGTGAAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGCTCGAA	245	
DB	195	CTGCATCAGTGAAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGCTCGAA	254	
QY	246	GGATGCAGGTTATGAGTAGCTCTGCATTTGATGACTGTTGGATGGCTCCCAAAGAGATTC	305	
DB	255	GGATGCAGGTTATGAGTAGCTCTGCATTTGATGACTGTTGGATGGCTCCCAAAGAGATTC	314	
QY	306	AGAAGGCAGACTTCAGGCAGACCTCAGCGCTTTCTCATGGGATTCGCCAGCTAGCTAA	365	
DB	315	AGAAGGCAGACTTCAGGCAGACCTCAGCGCTTTCTCATGGGATTCGCCAGCTAGCTAA	374	
QY	366	TTATGTTACAGCAAGGACTGAAGCTAGGGATTATGCGAGATTTGCGAAATAAAACCTG	425	
DB	375	TTATGTTACAGCAAGGACTGAAGCTAGGGATTATGCGAGATTTGCGAAATAAAACCTG	434	
QY	426	CGCAGGCTTCCCTGGGAGTTTGGATACTAGCAATTGATGCCCAGACCTTTGCTGACTG	485	
DB	435	CGCAGGCTTCCCTGGGAGTTTGGATACTAGCAATTGATGCCCAGACCTTTGCTGACTG	494	
QY	486	GGGAGTAGACTCTGCTAAAAATTGATGGTTGTTACTGTGACAGTTTGGAAAAATTGGCAGA	545	
DB	495	GGGAGTAGACTCTGCTAAAAATTGATGGTTGTTACTGTGACAGTTTGGAAAAATTGGCAGA	554	
QY	546	TGGTTATAAGCACAATGTCCTTGGCCCTGAATAGAGCTGGCAGAAAGCAATTGTGTACTCTG	605	
DB	555	TGGTTATAAGCACAATGTCCTTGGCCCTGAATAGAGCTGGCAGAAAGCAATTGTGTACTCTG	614	
QY	606	TGAGTGCCCTCTTATATGTGGCCCTTTCAAAGCCCCAAATTATACAGAAATCCGACAGTA	665	
DB	615	TGAGTGCCCTCTTATATGTGGCCCTTTCAAAGCCCCAAATTATACAGAAATCCGACAGTA	674	
QY	666	CTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAAAGTATAAAGAGTAT	725	
DB	675	CTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAAAGTATAAAGAGTAT	734	
QY	726	CTTGGACTGACATCTTTTAAACAGGAGAGAAATGTTGTATGTTGCTGGACAACA-GGGGGTT	784	
DB	735	CTTGGACTGACATCTTTTAAACAGGAGAGAAATGTTGTATGTTGCTGGACAACAAGGGGGTT	793	
QY	785	GGATGACCCAGGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCGCAAGTAA	844	
DB	794	GGATGACCCAGGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCGCAAGTAA	853	
QY	845	CTCAGATGGCCCTCTGGGCTATCATGGCTGCT-CCTTTTATTCATGCTCTAATGACCTCCGA	903	
DB	854	CTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCCTTTATTCATGCTCTAATGACCTCCG	913	
QY	904	CACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAGGATTAAGGACGTAAATTGCCATCAATCAG	963	
DB	914	AACATCAACCTC-AGCCAAAGCTCTCTTTCAAGATAAGGACGT-ATTGGCATTCATCAG	971	
QY	964	GACCCCTTGGCAGCAAGGGTA-CAAGCTTAGACAGGAGACAACCTTTGAAGTGTGGGA	1022	
DB	972	GAACCCCTGGNGCAGCCAGGGTACCCNNCTTAGACGCGAAACAACCTTTTAAATGTGGGA	1031	

Qy 1023 ACG-ACCTCTCTCAGGCTTAGCTGGG 1048
||| ||| ||| ||| ||| ||| ||| |||
Db 1032 ACGNACTTCTCTCAGGCTTTAAACGGG 1058

RESULT 10
AL577581/c
LOCUS
DEFINITION AL577581 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
CDNA clone CSODK009YE12 3-PRIME, mRNA sequence.
ACCESSION AL577581
KEYWORDS AL577581.3 GI:46256603
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1. (bases 1 to 1026)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)

On Feb 15, 2001 this sequence version replaced gi:31315840.

Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r

For more information about this cluster, see
<http://www.genoscope.cns.fr/cdna?s=CSODK009BC06NP1&c=10506.r>.

FEATURES
source
Location/Qualifiers
1..1026
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODK009YE12"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 72.3%; Score 915; DB 1; Length 1026;
Best Local Similarity 98.3%; Pred.No. 9.8e-263;
Matches 939; Conservative 4; Mismatches 10; Indels 2; Gaps 2;

Qy 289 GCTCCCCAAGAAGATTTCAGAGGCAGACTTCAGGCAGACCCTCAGCGCTTTCCTCATGGG 348
|||||
Db 955 GCTCCCCAAGAAGATTTCAGAA-CGAGACTTCA-GCAGACCCTCAGCGCTTTCCTCATGGG 898
|||||

Qy 349 ATTGCCAGCTAGCTAATTAATTGTTCACAGCAAAGGACTGAAGCTAGGGATTTATGCAGAT 408
|||||
Db 897 ATTCGCCAGCTAGCTAATTAATTGTTCACAGCAAAGGACTGAAGCTAGGGATTTATGCAGAT 838
|||||

Qy 409 GTTGGAATAAACCTCCGACGGCTTCCCTGGGAGTTTCGATACCTACGCATTTGATGCC 468
|||||
Db 837 GTTGGAATAAACCTCCGACGGCTTCCCTGGGAGTTTTCGATACCTACGCATTTGATGCC 778
|||||

Qy 469 CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGT 528
|||||
Db 777 CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGT 718
|||||

Qy 529 TTGAAAAATTGGCAGATGHTTTAAGCACATGTCTTTGGCCCTGTAATAGGACTGGCAGA 588
|||||
Db 717 TTGAAAAATTGGCAGATGHTTTAAGCACATGTCTTTGGCCCTGTAATAGGACTGGCAGA 658
|||||

QY	589	AGCATTGTCTACTCTCTGTGAGTGGCCCTCTTTATATGTGSCCCTTTCAAAAGCCCAATTAT	648	AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayaahizaki,Y.
Db	657	ACATTGTGTACTCTCTGTGAGTGGCCCTCTTTATATGTGSCCCTTTCAAAAGCCCAATTAT	598		RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer
QY	649	ACAGAAATCCGACAGTACTGCAATCACTGCGGAAATTTTGTGACATGATGATTCCTGG	708	TITLE	Genome Res. 10 (11), 1757-1771 (2000)
Db	597	ACAGAAATCCGACAGTACTGCAATCACTGCGGAAATTTTGTGACATGATGATTCCTGG	538	JOURNAL	4
QY	709	AAAAGTATAAAGAGTATCTTGACATGGACATCTTTTAACCAGAGAGAAATTTGATGTT	768	PUBMED	11076861
Db	537	AAAAGTATAAAGAGTATCTTGACATGGACATCTTTTAACCAGAGAGAAATTTGATGTT	478	REFERENCE	
QY	769	GCTGGACAGGGGGTGGAAATGACCCAGATATGTTAGTATGTCGCAACTTTGCGCTCAGC	828	AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Db	477	GCTGGACAGGGGGTGGAAATGACCCAGATATGTTAGTATGTCGCAACTTTGCGCTCAGC	418	JOURNAL	Functional annotation of a full-length mouse cDNA collection
QY	829	TGGAATCAGCAAGTAACTCAGATGGCCCTCTGCGCTATCATGGCTGCTCTTTATTCATG	888	AUTHORS	Nature 409, 695-690 (2001)
Db	417	TGGAATCAGCAAGTAACTCAGATGGCCCTCTGCGCTATCATGGCTGCTCTTTATTCATG	358	TITLE	5
QY	889	TCTATGACCTCGACACATCAGCCTCTAAGCCAAAGCTCTCTTCAGGATTAAGACGTA	948	JOURNAL	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Db	357	TCTATGACCTCGACACATCAGCCTCTAAGCCAAAGCTCTCTTCAGGATTAAGACGTA	298	REFERENCE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
QY	949	ATTGCCATCAATCAGGACCCCTTGGGCAAGCAAGGTACCAGCTTAGCAGGAGACAAC	1008	AUTHORS	Nature 420, 563-573 (2002)
Db	297	ATTGCCATCAATCAGGACCCCTTGGGCAAGCAAGGTACCAGCTTAGCAGGAGACAAC	238	JOURNAL	6 (bases 1 to 1376)
QY	1009	TTTGAAGTGGGAACGACCTCTCTCAGCGCTTAGCTGGGCTGTAGCTATGATAAACCGG	1068	REFERENCE	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayaehida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,Y., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
Db	237	TTTGAAGTGGGAACGACCTCTCTCAGCGCTTAGCTGGGCTGTAGCTATGATAAACCGG	178	TITLE	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
QY	1069	CAGGAGATTGGTGACCTCGCTTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTG	1128	JOURNAL	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken
Db	177	CAGGAGATTGGTGACCTCGCTTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTG	118	COMMENT	Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
QY	1129	GCTGTAATCCTGCGCTTTCATCACACAGCTCCTCCCTGTAAGGAAGTAGGTTTC	1188	FEATURES	Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/ Location/Qualifiers 1..1376 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:E330039P08" /db_xref="taxon:10090" /clone="E330039P08" /sex="female" /tissue type="ovary" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="2 days pregnant adult" 21..1286 /note="unnamed protein product; galactosidase, alpha (MGD MGJ:1347344, GB NM_013463, evidence: BLASTN, 99%, match=1373) /codon_start=1 /protein_id="BAC35819.1" /db_xref="GI:26344325" /translation="NAMKLLSRDTRLVCELALCPALVFWLSILGVRLDNGLRPTFM GWLHWFRCNLDCQEBPDACITSEQLFMQAEMLVSDGWRDAGYDLYLCIDCWAPER
Db	117	GCTGTAATCCTGCGCTTTCATCACACAGCTCCTCMTRTGAAGGAAGTAGGTTTC	58	source	
QY	1189	TATGAATGGACTTCAAGGTTAAGAAGTACATAAAATCCACAGGACCTGTTTTC	1243		
Db	57	TATGAATGGACTTCAAGGTTAAGAAGTACATAAAATCCACAGGACCTGTTTTC	3		
RESULT 11					
AK054547					
LOCUS		1376 bp	mRNA	linear	HTC 03-APR-2004
DEFINITION		Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN full-length enriched library, clone:E330039P08			
ACCESSION		AK054547			
VERSION		AK054547.1			
KEYWORDS		HTC; CAP trapper.			
SOURCE		Mus musculus (house mouse)			
ORGANISM		Mus musculus			
REFERENCE					
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
TITLE		1			
JOURNAL		Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
PUBMED		High-efficiency full-length cDNA cloning			
REFERENCE		Meth. Enzymol. 303, 19-44 (1999)			
AUTHORS		2			
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to Prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL		Genome Res. 10 (10), 1617-1630 (2000)			
PUBMED		11042159			
REFERENCE		3			


```

QY 1201 TCAGGTTAAGTACACATAAATCCACAGGCACTGTTTCTTCAGCTAGAAA 1255
Db 1207 TTGACCTTAAAAACGAGTAACCCCTCAGGCACTGTTTCTTCGTTAGAAA 1261

RESULT 13
CO645672/c
LOCUS
DEFINITION
ILLUMIGEN MCQ_25513 Katze MMLV Macaca mulatta cDNA clone
IBIUM:23923 5' similar to Bases 27 to 1049 highly similar to human
GLA (Hs.69089), mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Macaca mulatta (rhesus monkey)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Cercopithecoidea; Cercopithecinae; Macaca.
1 (bases 1 to 1133)
Magnes, C.B., Fellin, P.C., Thomas, M.J., Korth, M.J., Agy, M.B.,
Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and
Iadonato, S.P.
Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human
Genome Biol. 6 (7), R60 (2005)
15998449
Contact: C. Magnes
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagnes@illumigen.com
Sequenced on 2004.03.05. 699 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCTCACTAAAGGGAACAAA
BACKWARD: CACTATAGGCGCAATTGGTA
Insert Length: 1133 Std Error: 0.00
Plate: CL000111 Row: E Column: 08
Seq primer: CCTCACTAAAGGGAACAAA
POLYA=No.

FEATURES
Location/Qualifiers
1..1133
/organism="Macaca mulatta"
/mol_type="mRNA"
/strains="Indian"
/db_xref="taxon:9544"
/clone="IBIUM:23923"
/sex="female"
/dev_stage="adult"
/lab_host="Electronax DH10B"
/clone_lib="katze_MMLV"
/note="Organ: liver; Vector: pDONR 222; Site 1: BsrG I;
Site 2: BsrG I; Created from CloneMiner cDNA Library
Construction kit (catalog #18249-029)"

ORIGIN
Query Match 70.3%; Score 890; DB 7; Length 1133;
Best Local Similarity 94.3%; Pred. No. 3.3e-255;
Matches 945; Conservative 0; Mismatches 55; Indels 2; Gaps 2;
QY 264 CCTCTGCATTGATGACTGTTGGATGCTCCCAAGAGATTCCAGAGGCACTTCAGGC 323
Db 1041 CCTCTGCCCCCGGACTTTTGGAAATTCCTCCAAAGAGACTCGAAGCGCGCTTCGGGC 982
QY 324 AGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTAGCTAATTATGTTTCACAGCAAGG 383
Db 981 GGGCGTTAGTCTTCTTCATGGGATTCGCCAGTAGCTAATTATGATACAGCAAGG 922
QY 384 ACTGAAGCTAGGATTTATGACAGATGTTGGAAATAAAACCTCGCAGGCTTCCTCGGAG 443

```

```

Db 921 AATGAAAATAGGATTTATGACAGGTTGGAAATAAAACCTCGCAGGCTTCCTGGGAG 862
QY 444 TTTTGGATACACACATTGATGCCAGACCTTTGCTGACTGGGAGTAGATCTCTAAA 503
Db 861 TTTTGGATACATGACATTGATGCCAGCTCTTGTGCTGACTGGGAGTAGATATGTGA 802
QY 504 ATTTGATGGTGTACTGCTGACAGTTTGGAAAAATTTGGCAGATGTTTATAGCACATGTC 563
Db 801 ATTTGATGGTGTACTGCTGACAGTTTGGAAAAATTTGGCAGATGTTTATAGCACATGTC 742
QY 564 CTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 623
Db 741 GTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 682
QY 624 GTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGACAGTACTGCAATCACTGGCGAAA 683
Db 681 GTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGACAGTACTGCAATCACTGGCGAAA 622
QY 684 TTTTGTGCTGACATGATGATTTCTCTGAAAAGTATAAAGAGTATCTTGGAGTGGACATCTTT 743
Db 621 TTTTGTGCTGACATGATGATTTCTCTGAAAAGTATAAAGAGTATCTTGGAGTGGACATCTTT 562
QY 744 TAACAGCAGAGAGAAATTTGTGATGTTGCTGGACCCAGGGGTTGGAATGACCCAGATATGT 803
Db 561 TAACAGCAGAGAGAAATTTGTGATGTTGCTGGACCCAGGGGTTGGAATGACCCAGATATGT 502
QY 804 AGTGATTTGGCAACTTTTGGCCCTCAGCTGGAATCAGCAAGTAACTCAGATGGCCCTCTGGGC 863
Db 501 AGTGATTTGGCAACTTTTGGCCCTCAGCTGGAATCAGCAAGTAACTCAGATGGCCCTCTGGGC 442
QY 864 TATCATGCTGCTCTCTTTATTCATGCTTAATGACCTCCGACACATCAGCCCTCAAGCCAA 923
Db 441 TATCATGCTGCTCTCTTTATTCATGCTTAATGACCTCCGACACATCAGCCCTCAAGCCAA 382
QY 924 AGCTCTCTCTCAGGATAGGAGCTAATTTGCCATCAATCAGAGCCCTTGGGCAAGCAAGG 983
Db 381 AGCTCTCTCTCAGGATAGGAGCTAATTTGCCATCAATCAGAGCCCTTGGGCAAGCAAGG 322
QY 984 GTACAGCTTTAGACAGGAGAGACAACTTTGAAGTGTGGAAACGACCTCTCTCAGGCTTAGC 1043
Db 321 GTACAGCTTTAGACAGGAGAGACAACTTTGAAGTGTGGAAACGACCTCTCTCAGGCTTAGC 262
QY 1044 CTGGGCTGTAGCTATGATAAAACCGCAGAGAGATGCTGAGACCTGCTCTTATACCATCGC 1103
Db 261 CTGGGCTGTAGCTATGATAAAACCGCAGAGAGATGCTGAGACCTGCTCTTATACCATCGC 202
QY 1104 AGTTGCTTCCCTGGTAAAGGAGTGGCTGTAAATCCTGCTGCTTCATCACACAGCTCCT 1163
Db 201 GGTTCCTTCCCTGGTAAAGGAGTGGCTGTAAATCCTGCTGCTTCATCACACAGCTCCT 142
QY 1164 CCTGTGAAAAGGAAAGCTAGGCTTCTATGAATGGACTTCAAGGTTAAGAAAGTCACATAAA 1223
Db 141 CCTGTG-GAAAAGGAAAGCTAGGCTTCTATGAATGGACTTCAAGGTTAAGAAAGTCACATAAA 83
QY 1224 TCCACAGGCACTGTTTGTCTTTCAGCTAGAGAAAACAATGTA 1265
Db 82 TCCACAGGCACTG-TTTGCTTGGCTAGAAAAATAAATGCA 42

RESULT 14
CO645623
LOCUS
DEFINITION
ILLUMIGEN MCQ_24642 Katze MMLV Macaca mulatta cDNA clone
IBIUM:23923 5' similar to Bases 5 to 920 highly similar to human
GLA (Hs.69089), mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Macaca mulatta (rhesus monkey)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

```

Cercopithecidae; Cercopithecinae; Macaca.
1 (bases 1 to 922)
Magnes, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agy, M.B.,
Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and
Iadonato, S.P.
Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human
Genome Biol. 6 (7), R60 (2005)
15998449
Contact: C. Magnes
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagnes@illumigen.com
Sequenced on 2004.03.02. 796 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCTCACTAAAGGACAAAA
BACKWARD: CACTATAGGCGAATTTGGTA
Insert Length: 922 Std Error: 0.00
Plate: CL000111 row: E column: 08
Seq primer: CCTCACTAAAGGACAAAA
POLYA=No. Location/Qualifiers
1. 922
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBIUV:23923"
/sex="female"
/dev_stage="adult"
/lab_host="Electromax DH10B"
/clone_lib="Katze_MMLV"
/notes="Organ: liver; Vector: pDONR 222; Site 1: BarG I;
Site 2: BarG I; Created from CloneMiner cDNA Library
Construction kit (catalog #18249-029)"

Query Match 69.8%; Score 883.2; DB 7; Length 922;
Best Local Similarity 98.0%; Pred. No. 3.3e-253;
Matches 894; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ATGACGTGAGGAAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCTGGCC 60
DB 11 ATGACGTGAGGAAACCCAGAAATGCATCTGGGCTGCGCGCTTGCCTTCCTGGCC 70
QY 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACATGGATGGCAAGGCGCT 120
DB 71 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACATGGATGGCAAGGCGCT 130
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 131 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 190
QY 181 GATTCTGCACTAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTGTCAGAAAGGC 240
DB 191 GATTCTGCACTAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTGTCAGATGGC 250
QY 241 TGAAGAGTGCAGGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 251 TGAAGAGTGCAGGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 310
QY 301 GATTTCAGAGGCGAGCTTCAGGCGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
DB 311 GATTTCAGAGGCGAGCTTCAGGCGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 370
QY 361 GCTAATTTATGTTTACAGCAAGGAGCTGAAGCTAGGAGTTATTCAGATGTTGAAATATA 420
DB 371 GCTAATTTATGTTTACAGCAAGGAGCTGAAGCTAGGAGTTATTCAGATGTTGAAATATA 430

QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTCGATACGACATTCATGATGCCAGACCTTTGCT 480
DB 431 ACCTGCGCAGGCTTCCCTGGGAGTTTTCGATACGACATTCATGATGCCAGACCTTTGCT 490
QY 481 GACTGGGAGTAGATCTGCTAAATAATTTGATGTTGTTTACTGTGACAGTTTGGAAATTTG 540
DB 491 GACTGGGAGTAGATCTGCTAAATAATTTGATGTTGTTTACTGTGACAGTTTGGAAATTTG 550
QY 541 GCAGATGTTTATAGCACATGCTCTTGGCCCTGATAGACATGGCAGAGCAATTTGTGTAC 600
DB 551 GCAGATGTTTATAGCACATGCTCTTGGCCCTGATAGACATGGCAGAGCAATTTGTGTAC 610
QY 601 TCCTGTGAGTGGCCCTTCTTATATGATGGCCCTTCAAAAGCCCAATTTACAGAAATCCGA 660
DB 611 TCCTGTGAGTGGCCCTTCTTATATGATGGCCATTTCAAAAGCCCAATTTACAGAAATCCGA 670
QY 661 CAGTACTGCAATCACTGCGGAAATTTTCTGACATTCATGATTCCTGAAAGATATAAG 720
DB 671 CAGTACTGCAATCACTGCGGAAATTTTCTGACATTCATGATTCCTGAAAGATATAAG 730
QY 721 AGTATCTTGGACTGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACACGG 780
DB 731 AGTATCTTGGACTGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACACGG 790
QY 781 GGTTCGAATGACCCAGATATGTTAGTGTGTTGCACTTTGGCCTCAGCTGGAATCAGCAA 840
DB 791 GGTTCGAATGACCCAGATATGTTAGTGTGTTGCACTTTGGCCTCAGCTGGAATCAGCAA 850
QY 841 GTAACCTAGATGGCCCTCTCGGCTATCATGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
DB 851 GTAACCTAGATGGCCCTCTCGGCTATCATGCTGCTCTCTTTATTCATGCTTAATGACCTC 910
QY 901 CGACATCATCAGC 912
DB 911 CGACATCATCAGC 922

RESULT 15
AL575861/c
LOCUS
DEFINITION
AL575861 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1067YJ01 3-PRIME, mRNA sequence.
ACCESSION
AL575861
VERSION
AL575861.3 GI:46248771
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 (bases 1 to 1071)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
On Feb 16, 2001 this sequence version replaced gi:31314157.
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, Cp 5706 - 91057 EVRY cedex - FRANCE
Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0D1067CE01NP1&c=10506.r.
Location/Qualifiers
1. 1071
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

FEATURES
source

/clone="CS0DI067YJ01"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match		69.6%;	Score 880.8;	DB 1;	Length 1071;
Best Local Similarity		91.6%;	Pred. No. 1.9e-252;		
Matches 985;		Conservative 20;	Mismatches 61;	Indels 9;	Gaps 8;
Qy	191	TCAGTGAGAAGCTCTTTCATGAGATGGCAGAGCTCATGCTCTCAGAA-GGCTGGAAGGAT	249		
Db	1071	TCAGTGAGAAGCTC-TCATGNANATGNCAGANCTCATGNTCTCAGAACGMCCTGAAAGNAT	1013		
Qy	250	GCAGGTTATGAGTACCTCTGC--ATTGATGACTGTTGGATGGCT-CCCCAAGAGATTCA	306		
Db	1012	GCAGTNTATGAGTACCTCTGCANTTTGATGACTGTTGNATGNTCTCCCCAAAGAGATNCA	953		
Qy	307	GAAGGCAGACTTTCAGSCAGACCTCAGCGCTTTTCTCATGGGATTCGCCAGCTAGCTAAT	366		
Db	952	GAAGNCACACTNCAGCAGACCTCAGCGCTTTTCTCATGGGATTCGCCAGCTAGCTAAT	893		
Qy	367	TATGTTTCACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTTGGAAATAAAACCTGC	426		
Db	892	TATGCTCACAGCAAGGACTGAAGCTA-GGATTTATGCAGATGTT-GAATAAAGACCTGC	835		
Qy	427	GCAGGCTTCCCTGGGAGTTTGGATACACTACGACATGATGCCAGACCTTTCCTGACTGG	486		
Db	834	GCAGGCTTCCCTGGGAGTTTGGATACACTACGACATGATGCCAGACCTTTCCTGACTGG	775		
Qy	487	CGAGTAGATCTGTAATAATTTGATGCTGCTACTGTCAGATTTGCAAAATTTGGCAGAT	546		
Db	774	GGAGTAGATCTGTAATAATTTGATGCTGCTACTGTCAGATTTGCAAAATTTGGCAGAT	715		
Qy	547	GGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTACTCTCTGT	606		
Db	714	GATATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTACTACCTAT	655		
Qy	607	GAGTGGCTCTTTATATGTTGGCCCTTTTCAAAAGCCCAAT-TATACAGAAATCCGACAGTA	665		
Db	654	VAGTRACCTCACTATATCTGCCCTTMYCAAAAGMCCAATMTATACAAAAATCCGACAGTA	595		
Qy	666	CTGCAATCACTGCGGAAATTTGCTGACATTCGATTCCTCGAAAAGTATAAAGATAT	725		
Db	594	CTGCAATMACTGGCAAAATTTAGCTGACATTCGATTCCTGAAAAGTATAAAGATAT	535		
Qy	726	CTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTTGCTGGACCAGGGGGTTG	785		
Db	534	CTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTTGCTGGACCAGGGGGTTG	475		
Qy	786	GAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAAGTAAC	845		
Db	474	GAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAASTMAC	415		
Qy	846	TCAGATGGCCCTCTGGGCTCATAGGCTGCTCTCTTTTATTCATGCTCTAATGACCTCCGACA	905		
Db	414	TCAGAT-GGCCCTCTGGGCTCATAGGCTGCTCTCTTTTATTCATGCTCTAATGACCTCCGACA	356		
Qy	906	CATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAGGACGTAATTTGCCATCAATCAGGA	965		
Db	355	CATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAGGACGTAATTTGCCATCAATCAGGA	296		
Qy	966	CCCTTTGGGCAAGCAAGGGTACAGCTTAGACAGGAGACAACCTTTGAAGTGTGGAAACG	1025		
Db	295	CCCTTTGGGCAAGCAAGGGTACCRGCTTAGACAGGAGACAACCTTTGAAGTGTGGAAACG	236		
Qy	1026	ACCTCTCTCAGGCTTAGGCTGGCTGTAGCTATGATAAACCGCAGGAGATTTGGTGACC	1085		
Db	235	ACVTCCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAACCGCAGGAGATTTGGTGACC	176		

Search completed: December 26, 2005, 08:29:58
Job time : 4823 secs

Qy	1086	TCGCTCTTATATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCCTGTAAATCCTGCTG	1145
Db	175	TCGCTCTTATATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCCTGTAAATCCTGCTG	116
Qy	1146	CTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACTTCAAG	1205
Db	115	CTTCATCACACAGCACCTCCCNKGTGAAAAGGAAGCTAGGGTTCTATGAATGGACNWTCAAG	56
Qy	1206	GTTAAGAAAGTCACATAAAATCCACAGGCACCTGTTTTGCTTCAGCTAGAAAAACACA	1260
Db	55	GTTAAGMAGTCACBTWAATCCCNACAGGCASGTGTTTGCKTCAGCTAGAAAAATACA	1

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1266	100.0	1266	3	US-09-993-059-11	Sequence 11, Appl
2	1266	100.0	1266	3	US-10-103-327-11	Sequence 11, Appl
3	1264	99.8	1264	3	US-09-993-059-13	Sequence 13, Appl
4	1264	99.8	1264	3	US-10-103-327-13	Sequence 13, Appl
5	1263.4	99.8	1278	3	US-09-993-059-7	Sequence 7, Appl
6	1263.4	99.8	1278	3	US-10-103-327-7	Sequence 7, Appl
7	1263.4	99.8	1290	3	US-09-993-059-3	Sequence 3, Appl
8	1263.4	99.8	1290	3	US-10-103-327-3	Sequence 3, Appl
9	1263.4	99.8	1296	3	US-09-993-059-9	Sequence 9, Appl
10	1263.4	99.8	1296	3	US-10-103-327-9	Sequence 9, Appl
11	1263.4	99.8	1308	3	US-09-993-059-5	Sequence 5, Appl
12	1263.4	99.8	1308	3	US-10-103-327-5	Sequence 5, Appl
13	1261.8	99.7	1343	3	US-08-928-881-18	Sequence 18, Appl
14	1261.8	99.7	1343	3	US-09-543-921-18	Sequence 18, Appl
15	1261.8	99.7	1343	3	US-09-266-014-3	Sequence 3, Appl
16	1261.8	99.7	1343	3	US-09-491-759-18	Sequence 18, Appl
17	1261.8	99.7	1393	2	US-07-602-824A-1	Sequence 1, Appl
18	1261.8	99.7	1393	2	US-07-983-457-1	Sequence 1, Appl
19	1261.8	99.7	1393	2	US-08-361-577-6	Sequence 6, Appl
20	1252.4	98.9	1254	3	US-09-993-059-15	Sequence 15, Appl
21	1252.4	98.9	1254	3	US-10-103-327-15	Sequence 15, Appl
22	1251	98.8	1272	3	US-09-993-059-17	Sequence 17, Appl
23	1251	98.8	1272	3	US-10-103-327-17	Sequence 17, Appl
24	1251	98.8	8234	3	US-09-626-127-14	Sequence 14, Appl

```
QY 301 GATTTCAGAGGAGACTTCAGGAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAGGAGACTTCAGGAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATATATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCGAGATTTGCGAAATAAA 420
Db 361 GCTAATATATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCGAGATTTGCGAAATAAA 420
QY 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACACGACATTCAGCAATTCATGCGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACACGACATTCAGCAATTCATGCGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTCAGATTTGCGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTCAGATTTGCGAAATTTG 540
QY 541 GCAGATGTTTATAAGCACATGCTTCCTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Db 541 GCAGATGTTTATAAGCACATGCTTCCTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
QY 601 TCCTGTAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGACATTCGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGACATTCGATGATTCCTGGAAAGTATAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGTTGATGTTGCTGCGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGTTGATGTTGCTGCGACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTCGAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATGTCGAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 841 GTAATCTAGATGGCCCTTCGGCTATCATGCTGCTCCTTTATTCATGCTCAATGACCTC 900
Db 841 GTAATCTAGATGGCCCTTCGGCTATCATGCTGCTCCTTTATTCATGCTCAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTACAGGAGAGACAACTTTGAAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTACAGGAGAGACAACTTTGAAAGTGG 1020
QY 1021 GAAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATATAACCGGCGAGAGATGGT 1080
Db 1021 GAAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATATAACCGGCGAGAGATGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
QY 1141 GCCTGTTTCATCAGCAGCTCCTCCTGTGAAAAAGAGCTAGGTTTCTATGAATGGACT 1200
Db 1141 GCCTGTTTCATCAGCAGCTCCTCCTGTGAAAAAGAGCTAGGTTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCCTCAGCTAGAAAAACA 1260
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCCTCAGCTAGAAAAACA 1260
QY 1261 ATGTAA 1266
Db 1261 ATGTAA 1266
```

RESULT 2

US-10-103-327-11

; Sequence 11, Application US/10103327

```
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
; US-10-103-327-11
```

```
Query Match 100.0%; Score 1266; DB 3; Length 1266;
Beat Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGCTTGGCTTCTGCTTCTGCCC 60
Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGCTTGGCTTCTGCTTCTGCCC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTTGGCAAGGCGCT 120
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTTGGCAAGGCGCT 120
QY 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCA 180
QY 181 GATTCTGTCATCAGTGAAGACTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTGTCATCAGTGAAGACTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTCATGATGATGCTTGGATGGCTCCCAAGA 300
Db 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTCATGATGATGCTTGGATGGCTCCCAAGA 300
QY 301 GATTTCAGAAAGGAGACTTCAGGAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAAAGGAGACTTCAGGAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATATATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCGAGATTTGCGAAATAAA 420
Db 361 GCTAATATATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCGAGATTTGCGAAATAAA 420
QY 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACACGACATTCAGCAATTCATGCGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACACGACATTCAGCAATTCATGCGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTCAGATTTGCGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTCAGATTTGCGAAATTTG 540
QY 541 GCAGATGTTTATAAGCACATGCTTCCTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Db 541 GCAGATGTTTATAAGCACATGCTTCCTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
QY 601 TCCTGTAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGACATTCGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGACATTCGATGATTCCTGGAAAGTATAAG 720
```

Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGTGATGATTCCTGGAAAGTATAAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTGCTGGACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTAGTATGTGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTAGTATGTGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCCCTTTATTCATCTCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCCCTTTATTCATCTCTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGCTTAATTTGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGCTTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGAGGGTACAGCTTAGACAGAGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGAGGGTACAGCTTAGACAGAGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGATTTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGATTTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTCTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTCTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTATCACAAGCTCCTCCTGTGAAAAGAGTGGCTTCTAATGAAATGGACT 1200
Db 1141 GCCTGCTTATCACAAGCTCCTCCTGTGAAAAGAGTGGCTTCTAATGAAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATTAATCCACAGGACTGTTTTCCTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAATTAATCCACAGGACTGTTTTCCTCAGCTAGAAAACACA 1260
Qy 1261 ATGTAA 1266
Db 1261 ATGTAA 1266

RESULT 3

US-09-993-059-13
; Sequence 13, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUNAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-09-993-059-13

Query Match 99.8%; Score 1264; DB 3; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTTCGCTTCTCGGCC 60
|||||

Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTTCGCTTCTCGGCC 60
Qy 61 CTGCTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Db 61 CTGCTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGGGCTTCATGTGCAACCTTTGACTGCGAGAGAGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGGGCTTCATGTGCAACCTTTGACTGCGAGAGAGCA 180
Qy 181 GATTCTGTCATCAGTGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
Db 181 GATTCTGTCATCAGTGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
Qy 241 TGGAGGATGAGGCTTATGACTACCTCTGCAATGATGACTGTTGGATGGCTCCCCAAGA 300
Db 241 TGGAGGATGAGGCTTATGACTACCTCTGCAATGATGACTGTTGGATGGCTCCCCAAGA 300
Qy 301 GATTGAGAGGAGCTTTCAGGAGAGCCCTCAGGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTGAGAGGAGCTTTCAGGAGAGCCCTCAGGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTTACAGCAAGAGCTGAAGCTAGGATTTATGAGATGTTGGAAATAA 420
Db 361 GCTAATTTATGTTACAGCAAGAGCTGAAGCTAGGATTTATGAGATGTTGGAAATAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAATAATTTGATGGTTTACTGTGACAGTTTGGAAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAATAATTTGATGGTTTACTGTGACAGTTTGGAAAAATTG 540
Qy 541 GCAGATGTTATAGACATCTCTTGTGGCCCTGTAAGAGCCCAATTTATACAGAAATCCGA 600
Db 541 GCAGATGTTATAGACATCTCTTGTGGCCCTGTAAGAGCCCAATTTATACAGAAATCCGA 600
Qy 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGCAATTTGATGATTCCTGGAAGTATAAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGCAATTTGATGATTCCTGGAAGTATAAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTAGTATGTGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTAGTATGTGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCCCTTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCCCTTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTCAGGATAAGGAGCTTAATTTGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTCAGGATAAGGAGCTTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGAGGGTACAGCTTAGACAGAGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGAGGGTACAGCTTAGACAGAGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGATTTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGATTTGGT 1080
Qy 1081 GGACCTCGCTTATACCATCGCAGTCTGCTTCCCTGGGTAAAGAGTGGCTTCTGTAATCCT 1140
Db 1081 GGACCTCGCTTATACCATCGCAGTCTGCTTCCCTGGGTAAAGAGTGGCTTCTGTAATCCT 1140


```
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-09-993-059-7

Query Match      99.8%; Score 1263.4; DB 3; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTCTGGCC 60
DB 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTCTGGCC 60
QY 61 CTGCTTTCTGGGACATCCCTGGGCTAGAGCACTGGAATGGATTGGCAAGCGCT 120
DB 61 CTGCTTTCTGGGACATCCCTGGGCTAGAGCACTGGAATGGATTGGCAAGCGCT 120
QY 121 ACCATGGGCTGGCTGACATGGGAGGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
DB 121 ACCATGGGCTGGCTGACATGGGAGGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
QY 181 GATTCCTGCATCAGTGAAGAGCTTTTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
DB 181 GATTCCTGCATCAGTGAAGAGCTTTTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
QY 241 TGGAGGATGAGGATTTATGAGTACCTCTGCAATGTATGACTGTGGATGGCTCCCAAGA 300
DB 241 TGGAGGATGAGGATTTATGAGTACCTCTGCAATGTATGACTGTGGATGGCTCCCAAGA 300
QY 301 GATTCAGAGGAGACTTCAGGAGACACCTTCAGGCTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAGAGGAGACTTCAGGAGACACCTTCAGGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTTATGTTACAGCAAGAGGACTGAAGCTTAGGATTTATGCAAGATGTTGGAAATAA 420
DB 361 GCTAATTTATGTTACAGCAAGAGGACTGAAGCTTAGGATTTATGCAAGATGTTGGAAATAA 420
QY 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATACGACATTCAGTCCAGACCTTTGCT 480
DB 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATACGACATTCAGTCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGAAGATTTGGAATAATTG 540
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGAAGATTTGGAATAATTG 540
QY 541 GCAGATGTTATAGCAGATGCTTTCCTGGGCTTGGATAGGCTTGGCAGAGCATTCGTAC 600
DB 541 GCAGATGTTATAGCAGATGCTTTCCTGGGCTTGGATAGGCTTGGCAGAGCATTCGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTTGGGCTTCAAAAGCCCAATATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTTGGGCTTCAAAAGCCCAATATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGGCAAAATTTGCTGACATTCATGATGATTCCTGGAAAGTATAAG 720
DB 661 CAGTACTGCAATCACTGGGCAAAATTTGCTGACATTCATGATGATTCCTGGAAAGTATAAG 720
QY 721 AGTATCTTGGATGAGATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
DB 721 AGTATCTTGGATGAGATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
QY 781 GGTGTGAATGACCCAGATATGTTAGTATTGGCAACTTTGGGCTCAGCTGGAATCAGCAA 840
DB 781 GGTGTGAATGACCCAGATATGTTAGTATTGGCAACTTTGGGCTCAGCTGGAATCAGCAA 840
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATCATGTCTAATGACCTC 900
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATCATGTCTAATGACCTC 900
```

```
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGATTAAGGACGTAATTCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGATTAAGGACGTAATTCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGGAGGTACAGCTTACAGGAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGGAGGTACAGCTTACAGGAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
QY 1081 GGACCTCGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAATCCT 1140
DB 1081 GGACCTCGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAATCCT 1140
QY 1141 GCCTGCTTATCAGACAGCTCTCTCTGTGTAAGGAGAGCTAGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTATCAGACAGCTCTCTCTGTGTAAGGAGAGCTAGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAAACA 1260
DB 1201 TCAAGGTTAAGAGTTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAAACA 1260
QY 1261 ATGTA 1265
DB 1261 ATGCA 1265
```

RESULT 6

US-10-103-327-7

; Sequence 7, Application US/10103327

; Patent No. 6890748

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; PRIOR FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 1278

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1275)

US-10-103-327-7

Query Match 99.8%; Score 1263.4; DB 3; Length 1278;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTCTGGCC 60

DB 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTCTGGCC 60

QY 61 CTGCTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAGGAGCGCT 120

DB 61 CTGCTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAGGAGCGCT 120

QY 121 ACCATGGGCTGGCTGACATGGGAGGCTTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180

DB 121 ACCATGGGCTGGCTGACATGGGAGGCTTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180

181 GATTCTGCAATCAGTGAGAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||
181 GATTCTGCAATCAGTGAGAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY
241 TGGAGGATGCAGGTATAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db |||||
241 TGGAGGATGCAGGTATAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
QY
301 GATTTCAGAGGAGAGCTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATCGCCAGCTA 360
Db |||||
301 GATTTCAGAGGAGAGCTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATCGCCAGCTA 360
QY
361 GCTAATTAATGTTTCACAGCAAGAGAGCTGAAGCTAGGATTTATGAGATGTTGGAATAAA 420
Db |||||
361 GCTAATTAATGTTTCACAGCAAGAGAGCTGAAGCTAGGATTTATGAGATGTTGGAATAAA 420
QY
421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACCTAGCAATGATGCCAGACCTTTGCT 480
Db |||||
421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACCTAGCAATGATGCCAGACCTTTGCT 480
QY
481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGAAAAATTG 540
Db |||||
481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGAAAAATTG 540
QY
541 GCAGATGGTTAAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
Db |||||
541 GCAGATGGTTAAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
QY
601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAGCCCAATATACAGAAATCCGA 660
Db |||||
601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAGCCCAATATATACAGAAATCCGA 660
QY
661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAAGTATAAG 720
Db |||||
661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAAGTATAAG 720
QY
721 AGTATCTTGGACTGGAACATCTTTTAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db |||||
721 AGTATCTTGGACTGGAACATCTTTTAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
QY
781 GGTGGATGACCCAGATATGTTAGTATGTCGCACTTTGGCTCAGCTGGAAATCAGCA 840
Db |||||
781 GGTGGAAATGACCCAGATATGTTAGTATGTTGCAACTTTTGGCTCAGCTGGAAATCAGCA 840
QY
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTATGCTAATGACCTC 900
Db |||||
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTATGCTAATGACCTC 900
QY
901 CGACACATCAGCCCTCAAGCCCAAGCTCTCCTTCAGGATPAAGGACGTAAATGCCATCAAT 960
Db |||||
901 CGACACATCAGCCCTCAAGCCCAAGCTCTCCTTCAGGATPAAGGACGTAAATGCCATCAAT 960
QY
961 CAGGACCCCTTGGGAGCAAGGATACAGCTTACAGGGAGACAACTTTGAAGTGG 1020
Db |||||
961 CAGGACCCCTTGGGAGCAAGGATACAGCTTACAGGGAGACAACTTTGAAGTGG 1020
QY
1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATCATATAA CCGGAGGAGATGGT 1080
Db |||||
1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATCATATAA CCGGAGGAGATGGT 1080
QY
1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCCTGTAATCCT 1140
Db |||||
1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCCTGTAATCCT 1140
QY
1141 GCCTGCTTATACACAGCTCCTCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db |||||
1141 GCCTGCTTATACACAGCTCCTCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY
1201 TCAAGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAAACA 1260
Db |||||
1201 TCAAGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAAACA 1260
QY
1261 ATGTA 1265

Db 1261 ATGCA 1265
RESULT 7
US-09-993-059-3
; Sequence 3, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monco H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-09-993-059-3
Query Match 99.8%; Score 1263.4; DB 3; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTCGCTTCCTGGCC 60
Db 1 ATGCAAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTCGCTTCCTGGCC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGACACATGGAAATGGCAAGAGCGCT 120
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGACACATGGAAATGGCAAGAGCGCT 120
QY 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCATGTGCAACCTTGACTGCCAGGAGAGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCATGTGCAACCTTGACTGCCAGGAGAGCA 180
QY 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAGAGGAGAGCTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAGAGGAGAGCTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTAATGTTTCACAGCAAGAGAGCTGAAGCTAGGATTTATGAGATGTTGGAATAAA 420
Db 361 GCTAATTAATGTTTCACAGCAAGAGAGCTGAAGCTAGGATTTATGAGATGTTGGAATAAA 420
QY 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACCTAGCAATGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACCTAGCAATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGAAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGAAAAATTG 540
QY 541 GCAGATGGTTAAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
Db 541 GCAGATGGTTAAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
QY 601 TCCTGTGAGTGGCCCTCTTTTATATGTCGCTTTCAAAGCCCAATATATACAGAAATCCGA 660

Db 601 TCCTGTGAGTGGCCCTCTTTATATGAGGCCCTTTCAAAAAGCCCAATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGGAAATTTTCTGACATTTGATGATTTCTGCGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGGAAATTTTCTGACATTTGATGATTTCTGCGAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGAGACATCTTTTAAACAGAGAGAAATTTGATGATTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGAGACATCTTTTAAACAGAGAGAAATTTGATGATTTGCTGGACAGGG 780
Qy 781 GGTGGATGACCCAGATATGTTAGTGAATGGCACTTTGGCCCTCAGCTGGAATCAGCA 840
Db 781 GGTGGATGACCCAGATATGTTAGTGAATGGCACTTTGGCCCTCAGCTGGAATCAGCA 840
Qy 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGTAAATGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGTAAATGGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGAGGTTACAGCTTAGACAGGGAGCAACTTTGAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGAGGTTACAGCTTAGACAGGGAGCAACTTTGAGTGTGG 1020
Qy 1021 GAAAGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGAGATGGT 1080
Db 1021 GAAAGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGAGATGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGACGTGCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGACGTGCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGAGAGCTAGGTTCTATGATGAGT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGAGAGCTAGGTTCTATGATGAGT 1200
Qy 1201 TCAAGTTTAAAGATCACATAAATCCACAGGACGTGTTTCTTCACTAGAAAACACA 1260
Db 1201 TCAAGTTTAAAGATCACATAAATCCACAGGACGTGTTTCTTCACTAGAAAACACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGTA 1265

RESULT 8

US-10-103-327-3
; Sequence 3, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-10-103-327-3

Query Match 99.8%; Score 1263.4; DB 3; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTCTTGGCC 60
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTCTTGGCC 60
Qy 61 CTCGTTCTCTGGGACATCCCTCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT 120
Db 61 CTCGTTCTCTGGGACATCCCTCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCAATGTGCAACCTTGTGCTGCGAAGAGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCAATGTGCAACCTTGTGCTGCGAAGAGCA 180
Qy 181 GATTCTCTGCATCAGTGAGAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 181 GATTCTCTGCATCAGTGAGAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Qy 241 TGGAGGATGAGGCTTATGATGATCTCTGCAATGATGATCTTGGATGGCTCCCAAGA 300
Db 241 TGGAGGATGAGGCTTATGATGATCTCTGCAATGATGATCTTGGATGGCTCCCAAGA 300
Qy 301 GATTTCAGAGGAGCTTTCAGGAGAGCCCTCAGGCTTTTCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAGGAGCTTTCAGGAGAGCCCTCAGGCTTTTCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTTTCAGCAAGAGCTGAGCTAGGATTTATGCAATGTTGGAATATAA 420
Db 361 GCTAATTTATGTTTCAGCAAGAGCTGAGCTAGGATTTATGCAATGTTGGAATATAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTTGATGCGCCAGACCTTGT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTTGATGCGCCAGACCTTGT 480
Qy 481 GACTGGGAGTGTAGTCTGTAATAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTG 540
Db 481 GACTGGGAGTGTAGTCTGTAATAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTG 540
Qy 541 GCAGATGGTTATAGCACTGCTTGGCCCTGTAATAGGACTGGCAGAGCAATTCGTGTAC 600
Db 541 GCAGATGGTTATAGCACTGCTTGGCCCTGTAATAGGACTGGCAGAGCAATTCGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAAAGCCCAATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAAAGCCCAATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGGAAATTTTCTGACATTTGATGATTTCTGCGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGGAAATTTTCTGACATTTGATGATTTCTGCGAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGAGACATCTTTTAAACAGAGAGAAATTTGATGATTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGAGACATCTTTTAAACAGAGAGAAATTTGATGATTTGCTGGACAGGG 780
Qy 781 GGTGGATGACCCAGATATGTTAGTGAATGGCACTTTGGCCCTCAGCTGGAATCAGCA 840
Db 781 GGTGGATGACCCAGATATGTTAGTGAATGGCACTTTGGCCCTCAGCTGGAATCAGCA 840
Qy 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGTAAATGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGTAAATGGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGAGGTTACAGCTTAGACAGGGAGCAACTTTGAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGAGGTTACAGCTTAGACAGGGAGCAACTTTGAGTGTGG 1020

```
QY 1021 GAAACACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
DB 1021 GAAACACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTCCTCCCTGGGTAAGAGAGTGGCCCTGTAAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGAGTTCCTCCCTGGGTAAGAGAGTGGCCCTGTAAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAGCTAGGCTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAGCTAGGCTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAAACA 1260
DB 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAAACA 1260
QY 1261 ATGTA 1265
DB 1261 ATGCA 1265

RESULT 9
US-09-993-059-9
; Sequence 9, Application US/09993059
; Patent No. 6897696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1296)
US-09-993-059-9

Query Match 99.8%; Score 1263.4; DB 3; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCCTGGCC 60
QY 61 CTCGTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCT 120
DB 61 CTCGTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCT 120
QY 121 ACATGGGTGGTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGGTGGTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGNAGC 240
DB 181 GATTCCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGNAGC 240
QY 241 TGGAGGATTCAGGTTATGAGTACCTCTGCATTTGATGACTGTTGATGGCTCCCAAGA 300
DB 241 TGGAGGATTCAGGTTATGAGTACCTCTGCATTTGATGACTGTTGATGGCTCCCAAGA 300
QY 301 GATTCAGAAAGGACATTCAGGCAGACCTTCAGCGCTTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAGAAAGGACATTCAGGCAGACCTTCAGCGCTTTTCCTCATGGGATTCGCCAGCTA 360
```

```
QY 361 GCTAATTATGTTTCACAGCAAGAGCTGAAGCTAGGATTTATGAGATGTTTGGAAATAAA 420
DB 361 GCTAATTATGTTTCACAGCAAGAGCTGAAGCTAGGATTTATGAGATGTTTGGAAATAAA 420
QY 421 ACCTGGCAGAGCTTCCCTGGGAGTTTGGATPACTACGACATTTGATGATGATGATGATGAT 480
DB 421 ACCTGGCAGAGCTTCCCTGGGAGTTTGGATPACTACGACATTTGATGATGATGATGATGAT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 540
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 541 GCAGATGTTTATAAGCAGATCTCCTTGGCCCTGAAATAGGACTGGCAGAAAGCATTTGTGATC 600
DB 541 GCAGATGTTTATAAGCAGATCTCCTTGGCCCTGAAATAGGACTGGCAGAAAGCATTTGTGATC 600
QY 601 TCCTGTGATGGCTCTTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
DB 601 TCCTGTGATGGCTCTTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATGATGATGATGATGATGATGAT 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATGATGATGATGATGATGATGAT 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGATGATGATGATGATGAT 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGATGATGATGATGATGAT 780
QY 781 GGTGGAATGACCCAGATATCTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 781 GGTGGAATGACCCAGATATCTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 840
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTCTTATTCATCTCTTAATGACCTC 900
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTCTTATTCATCTCTTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGCTAATTTGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGCTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
QY 1021 GAAACACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
DB 1021 GAAACACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTCCTCCCTGGGTAAGAGAGTGGCCCTGTAAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGAGTTCCTCCCTGGGTAAGAGAGTGGCCCTGTAAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAGCTAGGCTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAGCTAGGCTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAAACA 1260
DB 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAAACA 1260
QY 1261 ATGTA 1265
DB 1261 ATGCA 1265
```

```
RESULT 10
US-10-103-327-9
; Sequence 9, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
```

;; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
;; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
;; FILE REFERENCE: 008010087CPUS06
;; CURRENT APPLICATION NUMBER: US/10/103.327
;; CURRENT FILING DATE: 2002-03-20
;; PRIOR APPLICATION NUMBER: US/09/993.059
;; PRIOR FILING DATE: 2001-11-13
;; NUMBER OF SEQ ID NOS: 37
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 9
;; TYPE: DNA
;; LENGTH: 1296
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1296)
US-10-103-327-9

Query Match 99.8%; Score 1263.4; DB 3; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGGCTTGGCTTCTCTGGCC 60
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGGCTTGGCTTCTCTGGCC 60

Qy 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGAATGGATTGGCAAGAGCGCT 120
Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGAATGGATTGGCAAGAGCGCT 120

Qy 121 ACCATGGGCTGGCTGACCTGGAGGCGTTGATGCAACCTTGATGCGCAGGAGAGGCA 180
Db 121 ACCATGGGCTGGCTGACCTGGAGGCGTTGATGCAACCTTGATGCGCAGGAGAGGCA 180

Qy 181 GATTCTCTGCATCAGTGAAGCTTTCATGAGAGGCGAGAGCTCATGCTCTCAGAAGGC 240
Db 181 GATTCTCTGCATCAGTGAAGCTTTCATGAGAGGCGAGAGCTCATGCTCTCAGAAGGC 240

Qy 241 TGAAGGATGAGGTTATGAGTACCTCTGCATTGATGATGCTTGGATGGCTCCCAAGA 300
Db 241 TGAAGGATGAGGTTATGAGTACCTCTGCATTGATGATGCTTGGATGGCTCCCAAGA 300

Qy 301 GATTGAGAGGAGACTTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTGAGAGGAGACTTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

Qy 361 GCTAATATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAAGATTTGGAAATAA 420
Db 361 GCTAATATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAAGATTTGGAAATAA 420

Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACAGCAATGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACAGCAATGATGCCAGACCTTTGCT 480

Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTCGACAGTTTGGAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTCGACAGTTTGGAAAATTG 540

Qy 541 GCAGATGTTAAGCAGATGCTGCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
Db 541 GCAGATGTTAAGCAGATGCTGCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600

Qy 601 TCCTGTAGTGGCTCTTTATATGTTGGCTTTCAAAAGCCCAATATATACAGAAATCCGA 660
Db 601 TCCTGTAGTGGCTCTTTATATGTTGGCTTTCAAAAGCCCAATATATACAGAAATCCGA 660

Qy 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATGATGATTCCTGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATGATGATTCCTGAAAAGTATAAG 720

Qy 721 AGTATCTGGACTGACATCTTTTAAACGAGGAGAAATTTGATGTTGCTGGACACGG 780
Db 721 AGTATCTGGACTGACATCTTTTAAACGAGGAGAAATTTGATGTTGCTGGACACGG 780

Qy 781 GGTGGAAATGACCCAGATATGTTAGTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840

Qy 841 GTAACCTAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTTCATGCTCTAATGACCTC 900
Db 841 GTAACCTAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTTCATGCTCTAATGACCTC 900

Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGAGCTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGAGCTAATTTGCCATCAAT 960

Qy 961 CAGGACCCCTTGGGCAAGAGGCTACAGCTTACAGGAGAGCAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGAGGCTACAGCTTACAGGAGAGCAACTTTGAAGTGTGG 1020

Qy 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080

Qy 1081 GGACCTCGCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGGAGTGGCCTGTAACTCT 1140
Db 1081 GGACCTCGCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGGAGTGGCCTGTAACTCT 1140

Qy 1141 GCCTGCTTATCAGCAGCTCTCTGCTGTAAGAGGAGCTAGGTTCTATGAAATGGACT 1200
Db 1141 GCCTGCTTATCAGCAGCTCTCTGCTGTAAGAGGAGCTAGGTTCTATGAAATGGACT 1200

Qy 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGAGCTGTTTGTCTTCACTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGAGCTGTTTGTCTTCACTAGAAAACACA 1260

Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 11

US-09-993-059-5
; Sequence 5, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993.059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1308)
US-09-993-059-5

Query Match 99.8%; Score 1263.4; DB 3; Length 1308;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGCCTTGGCTTCTCTGGCC 60
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGCCTTCTCTGGCC 60

Qy 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGGATTGGCAAGGAGCGCT 120
Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGGATTGGCAAGGAGCGCT 120

QY 721 ACTATCTTGGA CTGGACATCTTTTAAACAGGAGAGAA TTGTTGATGTTGCTCGACAGGG 780
Db 756 AGTATCTTGGA CTGGACATCTTTTAAACAGGAGAGAA TTGTTGATGTTGCTCGACAGGG 815
QY 781 GGTGGAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGGCTCAGCTGGAATCAGCAA 840
Db 816 GGTGGAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGGCTCAGCTGGAATCAGCAA 875
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGTGCTCTCTTTTATCATGCTTAATGACCTC 900
Db 876 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGTGCTCTCTTTTATCATGCTTAATGACCTC 935
QY 901 CCACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGTAATTCGCATCAAT 960
Db 936 CCACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGTAATTCGCATCAAT 995
QY 961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTTAGACAGGAGAGACAATTTGAAAGTGTGG 1020
Db 996 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTTAGACAGGAGAGACAATTTGAAAGTGTGG 1055
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGAGAGAA TTGGT 1080
Db 1056 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGAGAGAA TTGGT 1115
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
Db 1116 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1175
QY 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAGGAAGCTAGGTTTCTATGAATGGACT 1200
Db 1176 GCCTGCTTCATCACAGCTCCTCCTGTGAAAGGAAGCTAGGTTTCTATGAATGGACT 1235
QY 1201 TCAAGGTTAAGAGTCACATAATCCACAGGACGTGTTTTCCTTCAGCTAGAAAAACA 1260
Db 1236 TCAAGGTTAAGAGTCACATAATCCACAGGACGTGTTTTCCTTCAGCTAGAAAAACA 1295
QY 1261 ATGTA 1265
Db 1296 ATGCA 1300

RESULT 14
US-09-543-921-18
; Sequence 18, Application US/09543921
; Patent No. 6395884
; GENERAL INFORMATION:
; APPLICANT: Selden et al., Richard F.
; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A
; DEFICIENCY
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/543,921
; FILING DATE: 06-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,881
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819

; REFERENCE/DOCKET NUMBER: 07236/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1343 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-543-921-18

Query Match 99.7%; Score 1261.8; DB 3; Length 1343;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1263; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGCAAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTGGCGCTTCGCTTCTCGGCC 60
Db 36 ATGCAAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTGGCGCTTCGCTTCTCGGCC 95
QY 61 CTCGTTTCTGGAACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGAGCGCT 120
Db 96 CTCGTTTCTGGAACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGAGCGCT 155
QY 121 ACCATGGGCTGGCTGACCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 180
Db 156 ACCATGGGCTGGCTGACCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 215
QY 181 GATTCCTGCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 216 GATTCCTGCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 275
QY 241 TGGAAAGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTGTGATGGCTCCCCAAGA 300
Db 276 TGGAAAGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTGTGATGGCTCCCCAAGA 335
QY 301 GATTCAGAAAGCAGACTTTCAGGAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 336 GATTCAGAAAGCAGACTTTCAGGAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 395
QY 361 GCTAATTATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAAGATGTTGGAATAAA 420
Db 396 GCTAATTATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAAGATGTTGGAATAAA 455
QY 421 ACCTCGGACAGGCTTCCCTGGAGTTTGGATACATAGACATTTGATGCCAGACCTTTGCT 480
Db 456 ACCTCGGACAGGCTTCCCTGGAGTTTGGATACATAGACATTTGATGCCAGACCTTTGCT 515
QY 481 GACTGGGAGTAGATCTGTAATAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTG 540
Db 516 GACTGGGAGTAGATCTGTAATAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTG 575
QY 541 GCAGATGGTTATAAGCACATGCTTCCTGGCCCTGAAATAGGACTGGCAGAGCAATGTGTAC 600
Db 576 GCAGATGGTTATAAGCACATGCTTCCTGGCCCTGAAATAGGACTGGCAGAGCAATGTGTAC 635
QY 601 TCCTGTGATGGCCCTCTTTATATGATGGCCCTTTTCAAAAGCCCAATATACAGAAATCCGA 660
Db 636 TCCTGTGATGGCCCTCTTTATATGATGGCCCTTTTCAAAAGCCCAATATACAGAAATCCGA 695
QY 661 CAGTACTGCAATCAGTGGGAAATTTTCTGCACTTGATGATTCCTGGAAAAGTATAAG 720
Db 696 CAGTACTGCAATCAGTGGGAAATTTTCTGCACTTGATGATTCCTGGAAAAGTATAAG 755
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACAGGG 780
Db 756 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACAGGG 815
QY 781 GGTGGAATGACCCAGATATGTTAGTGATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
Db 816 GGTGGAATGACCCAGATATGTTAGTGATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 875

```
Qy 841 GTAACATGAGGCGCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTCTAATGACCTC 900
Db 876 GTAACATGAGGCGCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTCTAATGACCTC 935
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGTAAATGGCCATCAAT 960
Db 936 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGTAAATGGCCATCAAT 995
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTPACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 996 CAGGACCCCTTGGGCAAGCAAGGTPACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1055
Qy 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Db 1056 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1115
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db 1116 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1175
Qy 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCATGAATGGACT 1200
Db 1176 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCATGAATGGACT 1235
Qy 1201 TCAGGTTAAGAGTACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAAACACA 1260
Db 1236 TCAGGTTAAGAGTACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAAATACA 1295
Qy 1261 ATGTA 1265
Db 1296 ATGCA 1300
```

RESULT 15

US-09-266-014-3

; Sequence 3, Application US/09266014

; Patent No. 6458574

; GENERAL INFORMATION:

; APPLICANT: Selden, Richard F

; APPLICANT: Borowski, Marianne

; APPLICANT: Kinoshita, Carol M

; APPLICANT: Treco, Douglas A

; APPLICANT: Williams, Melanie D

; APPLICANT: Schuetz, Thomas J

; APPLICANT: Daniel, Peter F.

; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency

; FILE REFERENCE: FABRY DISEASE (18082-001)

; CURRENT APPLICATION NUMBER: US/09/266,014

; PRIOR FILING DATE: 1999-03-11

; PRIOR APPLICATION NUMBER: 60/026,041

; PRIOR FILING DATE: 1996-09-13

; PRIOR APPLICATION NUMBER: 08/928,881

; PRIOR FILING DATE: 1997-09-12

; PRIOR APPLICATION NUMBER: PCT US97/16603

; PRIOR FILING DATE: 1997-09-12

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 1343

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-266-014-3

Query Match 99.7%; Score 1261.8; DB 3; Length 1343;

Best Local Similarity 99.8%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;
Matches 1263; Conservative 0;

```
Qy 1 ATGACGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTTGGCGCTTCGCTTCCTGGCC 60
Db 36 ATGACGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTTGGCGCTTCGCTTCCTGGCC 95
Qy 61 CTCGCTTCTGGAATCCCTCGGCTAGAGCACTGAGATGGATGGCAAGGACGCT 120
```

```
Db 96 CTCGTTTCTGGGACATCCCTGGGCTTAGAGCACTGGACATGATTTGGCAAGGACGCT 155
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCGCAGGAAGAGCA 180
Db 156 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCGCAGGAAGAGCA 215
Qy 181 GATTTCCTGTCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
Db 216 GATTTCCTGTCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 275
Qy 241 TGGAAAGATGAGGCTTATGAGTACTCTTGCAATTCATGACTCTTGGAATGGCTCCCAAGA 300
Db 276 TGGAAAGATGAGGCTTATGAGTACTCTTGCAATTCATGACTCTTGGAATGGCTCCCAAGA 335
Qy 301 GATTTCAGAGGAGAGCTTCAGGAGAGAGCTTCAGGCTTTTCTCATGGAATTCGCCAGCTA 360
Db 336 GATTTCAGAGGAGAGCTTCAGGAGAGAGCTTCAGGCTTTTCTCATGGAATTCGCCAGCTA 395
Qy 361 GCTAATTTATGTTTACAGCAAGGACTGAAGCTAGGATTTATGAGATGTTTGGAAATAAA 420
Db 396 GCTAATTTATGTTTACAGCAAGGACTGAAGCTAGGATTTATGAGATGTTTGGAAATAAA 455
Qy 421 ACCTGGCAGGCTTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 480
Db 456 ACCTGGCAGGCTTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 515
Qy 481 GACTGGGAGTGAATCTGCTAAAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAAATTG 540
Db 516 GACTGGGAGTGAATCTGCTAAAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAAATTG 575
Qy 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAGCAATTTGTGTAC 600
Db 576 GCAGATGGTTATAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAGCAATTTGTGTAC 635
Qy 601 TCCTGTGAGTGGCTCTTTTATATGTTGGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 660
Db 636 TCCTGTGAGTGGCTCTTTTATATGTTGGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 695
Qy 661 CAGTACTGCAATCACTGCGGAAATTTTGTGCAATTTGATGATTTCTGGAAGATATAAG 720
Db 696 CAGTACTGCAATCACTGCGGAAATTTTGTGCAATTTGATGATTTCTGGAAGATATAAG 755
Qy 721 AGTATCTTGGATGGACATCTTTTAAACAGAGAGATTTGTTGATTTGCTGGACAGGG 780
Db 756 AGTATCTTGGATGGACATCTTTTAAACAGAGAGATTTGTTGATTTGCTGGACAGGG 815
Qy 781 GGTGGAATGACCCAGAGATATGTTAGTATTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Db 816 GGTGGAATGACCCAGAGATATGTTAGTATTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 875
Qy 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTTAATGACCTC 900
Db 876 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTTAATGACCTC 935
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGAGCTTAATTTGCCATCAAT 960
Db 936 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGAGCTTAATTTGCCATCAAT 995
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 996 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1055
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Db 1056 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1115
Qy 1081 GGACCTGCTCTTATACATCGAGTGTCTTCCCTGGGTAAGGAGTGGCGCTTGTAAATCCT 1140
Db 1116 GGACCTGCTCTTATACATCGAGTGTCTTCCCTGGGTAAGGAGTGGCGCTTGTAAATCCT 1175
Qy 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTAAGGAGAGCTAGGCTTCTATGAATGGACT 1200
Db 1176 GCCTGCTTCATCACACAGCTCTCTCCCTGTAAGGAGAGCTAGGCTTCTATGAATGGACT 1235
```

Qy 1201 TCAAGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTAGAAAACACA 1260
Db |||||
Qy 1236 TCAAGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTAGAAAATACA 1295
Db |||||

Qy 1261 ATGTA 1265
Db ||||
1296 ATGCA 1300

Search completed: December 26, 2005, 08:34:32
Job time : 268 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 07:09:39 ; Search time 1004 Seconds
(without alignments)
10427.323 Million cell updates/sec

Title: US-10-602-219-11
Perfect score: 1266
Sequence: 1 atcgagctgaggaacccaga.....agctagaacacaaatgtaa 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 993542 seqs, 413469005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA_Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	100.0	1266	3	US-09-993-059-11
2	1266	100.0	1266	5	US-10-103-327-11
3	1266	100.0	1266	7	US-10-602-219-11
4	1266	100.0	1266	7	US-10-602-220-11
5	1266	100.0	1266	8	US-10-851-388-11
6	1266	100.0	1266	9	US-10-984-389-11
7	1264	99.8	1284	5	US-09-993-059-13
8	1264	99.8	1284	5	US-10-103-327-13
9	1264	99.8	1284	7	US-10-602-219-13
10	1264	99.8	1284	7	US-10-602-220-13
11	1264	99.8	1284	8	US-10-851-388-13
12	1264	99.8	1284	9	US-10-984-389-13
13	1263.4	99.8	1278	3	US-09-993-059-7
14	1263.4	99.8	1278	5	US-10-103-327-7
15	1263.4	99.8	1278	7	US-10-602-219-7
16	1263.4	99.8	1278	7	US-10-602-220-7
17	1263.4	99.8	1278	8	US-10-851-388-7
18	1263.4	99.8	1278	9	US-10-984-389-7
19	1263.4	99.8	1290	5	US-09-993-059-3
20	1263.4	99.8	1290	5	US-10-103-327-3
21	1263.4	99.8	1290	7	US-10-602-219-3
22	1263.4	99.8	1290	7	US-10-602-220-3
23	1263.4	99.8	1290	8	US-10-851-388-3

ALIGNMENTS

RESULT 1
US-09-993-059-11
; Sequence 11, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-993-059-11

Query Match	100.0%	Score 1266;	DB 3;	Length 1266;
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 0;	Indels 0;
Matches 1266;	Conservative	0;		Gaps 0;
Qy	1	ATGCAGCTGAGGAACCCAGAACACTACATCTGGGCTGCGGCTTGGCGCTTCGCTTCCTGGCC	60	
Db	1	ATGCAGCTGAGGAACCCAGAACACTACATCTGGGCTGCGGCTTGGCGCTTCGCTTCCTGGCC	60	
Qy	61	CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT	120	
Db	61	CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT	120	
Qy	121	ACCATGGGCTGGCTGCACTGGGAGCGCTTCACTGCAACCTTGAATGCAAGAGGCA	180	
Db	121	ACCATGGGCTGGCTGCACTGGGAGCGCTTCACTGCAACCTTGAATGCAAGAGGCA	180	
Qy	181	GATTCTCTGCATCAGTGAAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240	
Db	181	GATTCTCTGCATCAGTGAAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240	
Qy	241	TGGAAGGATGAGGTTATGATACCTCTGCAATGATGATGGCTCCCAAGA	300	
Db	241	TGGAAGGATGAGGTTATGATACCTCTGCAATGATGATGGCTCCCAAGA	300	

```
Db 241 TGGAAAGATGCAAGGTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAAGA 300
Qy 301 GATTTCAGAGGAGACTTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAGGAGAGACTTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATATATGTTTACAGCAAAAGACTGAAGCTAGGATTTATGCAAGATGTTGGAAATAAA 420
Db 361 GCTAATATATGTTTACAGCAAAAGACTGAAGCTAGGATTTATGCAAGATGTTGGAAATAAA 420
Qy 421 ACCTCGCAGGCTTCCTCGGAGTTTTGGATACTAGCAATGATGCCAGACCTTTGCT 480
Db 421 ACCTCGCAGGCTTCCTCGGAGTTTTGGATACTAGCAATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTTGTACTGTGACAGATTTGGAAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTTGTACTGTGACAGATTTGGAAAAATTG 540
Qy 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Db 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATGTTGGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGTTGGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCAGTCGGGAAATTTGCTGCATTTGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCAGTCGGGAAATTTGCTGCATTTGATGATTCCTGGAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGACATCTTTTAAACAGAGAGAAATTTGTTGATGTTGCTGGACACAGG 780
Db 721 AGTATCTTGGACTGACATCTTTTAAACAGAGAGAAATTTGTTGATGTTGCTGGACACAGG 780
Qy 781 GGTTCGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Db 781 GGTTCGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATCATGTCATAGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATCATGTCATAGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGTAAATTCGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGTAAATTCGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTTAGACAGGGAGACAACCTTTGAAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTTAGACAGGGAGACAACCTTTGAAAGTGG 1020
Qy 1021 GAACGACCTCTCAGGCTTAGCCCTGAGCTGATGATGATAAACGGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCAGGCTTAGCCCTGAGCTGATGATGATAAACGGCAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAGATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAGATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAAACACA 1260
Qy 1261 ATGTAA 1266
Db 1261 ATGTAA 1266
```

RESULT 2

US-10-103-327-11

```
; Sequence 11, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CUPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-103-327-11
```

```
Query Match 100.0%; Score 1266; DB 5; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGCTTGGCTTGGCTTCTCTGGCC 60
Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTGGCTTCTCTGGCC 60
Qy 61 CTCGTTTCTTGGEACATCCCTGGGCTAGAGCACTGGCAATGGATGGCAAGGACGCT 120
Db 61 CTCGTTTCTTGGEACATCCCTGGGCTAGAGCACTGGCAATGGATGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGCTCAGCTGGGAGGCTTCATGTGCAACCTTGGCTGCGCAGGAGGCA 180
Db 121 ACCATGGGCTGCTGCACTGGGAGGCTTCATGTGCAACCTTGGCTGCGCAGGAGGCA 180
Qy 181 GATTCTTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 181 GATTCTTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Qy 241 TGGAAAGATGCAAGTTATGATACCTCTGCATTTGATGATGATGGATGGCTCCCAAGA 300
Db 241 TGGAAAGATGCAAGTTATGATACCTCTGCATTTGATGATGATGGATGGCTCCCAAGA 300
Qy 301 GATTTCAGAAAGCAGACTTCAGGAGACCTTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAAAGCAGACTTCAGGAGACCTTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATATATGTTTACAGCAAAAGACTGAAGCTAGGATTTATGCAAGATGTTGGAAATAAA 420
Db 361 GCTAATATATGTTTACAGCAAAAGACTGAAGCTAGGATTTATGCAAGATGTTGGAAATAAA 420
Qy 421 ACCTCGCAGGCTTCCTCGGAGTTTTGGATACTAGCAATGATGCCAGACCTTTGCT 480
Db 421 ACCTCGCAGGCTTCCTCGGAGTTTTGGATACTAGCAATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTTGTACTGTGACAGATTTGGAAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTTGTACTGTGACAGATTTGGAAAAATTG 540
Qy 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Db 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATGTTGGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGTTGGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCAGTCGGGAAATTTGCTGCATTTGATGATTCCTGGAAAGTATAAG 720
```

Db 661 CAGTACTGCAATCACTGGGAAATTTGCTGACATTTGATTTCTGGAAAGATATAAG 720
Qy 721 AGTATCTTGGACTGACATCTTTTAAACAGGAGAGAAATTTGTTGATTTGCTGGACACGG 780
Db 721 AGTATCTTGGACTGACATCTTTTAAACAGGAGAGAAATTTGTTGATTTGCTGGACACGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGTCGAACTTTGGCCTCAGCTGGAATCAGAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTCGAACTTTGGCCTCAGCTGGAATCAGAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTCATCTCTAAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTCATCTCTAAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGCTTAATGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGCTTAATGGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCAGGCTTAGCCTGAGCTGCTGGCTGTAGCTATGATAAACCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCAGGCTTAGCCTGAGCTGCTGGCTGTAGCTATGATAAACCAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAAAGCTAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAAAGCTAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGCTTAAGAACTACATAAATCCACAGGACCTGTTTGGCTCAGCTAGAAAACACA 1260
Db 1201 TCAAGCTTAAGAACTACATAAATCCACAGGACCTGTTTGGCTCAGCTAGAAAACACA 1260
Qy 1261 ATGTAA 1266
Db 1261 ATGTAA 1266

RESULT 3
US-10-602-219-11
; Sequence 11, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692

; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-11

Query Match 100.0%; Score 1266; DB 7; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTGGCC 60
Qy 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTTGCAAGGACGCT 120
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTTGCAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGCA 180
Qy 181 GATTCTGTCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTGTCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGGAGGATGAGCTTATGAGTACCTCTGCAATTTGATGATCTGTCATGGCTCCCAAGA 300
Db 241 TGGAGGATGAGCTTATGAGTACCTCTGCAATTTGATGATCTGTCATGGCTCCCAAGA 300
Qy 301 GATTCAAGAGCAGACTTTCAGGAGACCTTCAGGAGCTTCAGGCTTTCCTCATGGGATTC 360
Db 301 GATTCAAGAGCAGACTTTCAGGAGACCTTCAGGAGCTTCAGGCTTTCCTCATGGGATTC 360
Qy 361 GCTAATTTATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAAGATTTGCAATATA 420
Db 361 GCTAATTTATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAAGATTTGCAATATA 420
Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTTTCGATACATGATGATGATGATGATGATGATGAT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTTCGATACATGATGATGATGATGATGATGATGAT 480
Qy 481 GACTGGGAGTAGATCTCTAAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTCTAAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 541 CGAGATGTTTAAAGCACAATGCTTGGCCCTGAAATAGGACTGGCAGAGCAATTTGTGTAC 600
Db 541 CGAGATGTTTAAAGCACAATGCTTGGCCCTGAAATAGGACTGGCAGAGCAATTTGTGTAC 600
Qy 601 TCCTGTGATGGCTCTTTTATATGTCGCTTTCAGAGCCCAATATATACAGAAATCCGA 660
Db 601 TCCTGTGATGGCTCTTTTATATGTCGCTTTCAGAGCCCAATATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTGGAAAGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTGGAAAGATATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTGCTGGACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGTCGAACTTTGGCTCAGCTGGAATCAGCA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTCGAACTTTGGCTCAGCTGGAATCAGCA 840


```
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGGCTTCGCTTCCTGGCC 60
Qy 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGATGGCAAGGAGCGCT 120
Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGATGGCAAGGAGCGCT 120
Qy 121 ACCATGGGCTGGCTGCTGCTAGAGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
Db 121 ACCATGGGCTGGCTGCTGCTAGAGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
Qy 181 GATTCCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGC 240
Db 181 GATTCCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGC 240
Qy 241 TGGAAAGGATGCAAGGTATGAGTACCTCTGCAATGTGATGACTGTGGATGGCTCCCAAGA 300
Db 241 TGGAAAGGATGCAAGGTATGAGTACCTCTGCAATGTGATGACTGTGGATGGCTCCCAAGA 300
Qy 301 GATTCAGAGGAGCACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAGAGGAGCACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATATGTTTCAAGCAAGGACTGAGCTAGGATTTATGCAAGATGTTGGAATAAA 420
Db 361 GCTAATATGTTTCAAGCAAGGACTGAGCTAGGATTTATGCAAGATGTTGGAATAAA 420
Qy 421 ACCTCGGAGGCTTCCTGGGAGTTTGGATACAGCACTGAGCTAGGATTTATGCAAGATGTTG 480
Db 421 ACCTCGGAGGCTTCCTGGGAGTTTGGATACAGCACTGAGCTAGGATTTATGCAAGATGTTG 480
Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGATTTGGAATAATG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGATTTGGAATAATG 540
Qy 541 GCAGATGTTTAAAGCAATGCTTGGCCCTGAATAGCACTGGCAGAGCAATGTTGTAC 600
Db 541 GCAGATGTTTAAAGCAATGCTTGGCCCTGAATAGCACTGGCAGAGCAATGTTGTAC 600
Qy 601 TCCTGTGAGTGCCCTCTTTATATGTGGCCCTTCAAAGCCCAATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGCCCTCTTTATATGTGGCCCTTCAAAGCCCAATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATGATGATTCCTGGAAAGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATGATGATTCCTGGAAAGATATAAG 720
Qy 721 AGTATCTGGACTGGACATCTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTGGACTGGACATCTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Qy 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGNAATCAGAA 840
Db 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGNAATCAGAA 840
Qy 841 GTAACCTCAGATGGCCTCTGGGCTATCATGGCTGCTCTTTTATCATGTCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCTCTGGGCTATCATGGCTGCTCTTTTATCATGTCTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTAATGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTAATGGCCATCAAT 960
Qy 961 CAGGACCCCTGGCAGCAAGGATACAGCTTAGACAGGAGACAACTTGAAGTGTGG 1020
Db 961 CAGGACCCCTGGCAGCAAGGATACAGCTTAGACAGGAGACAACTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACGGCAGGAGATGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACGGCAGGAGATGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATGCAAGTTCCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATGCAAGTTCCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
```

```
Db 1081 GGACCTCGCTCTTATACCATGCAAGTTCCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAAGAAAGCTAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAAGAAAGCTAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTTCACATAAATCCCAAGGCACTGTTTTCCTCAGCTAGAAACACA 1260
Db 1201 TCAAGGTTAAGAGTTCACATAAATCCCAAGGCACTGTTTTCCTCAGCTAGAAACACA 1260
Qy 1261 ATGT 1264
Db 1261 ATGT 1264
```

RESULT 8

```
US-10-103-327-13
; Sequence 13, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103.327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICANT NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-103-327-13
```

Query Match 99.8%; Score 1264; DB 5; Length 1284;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGGCTTCGCTTCCTGGCC 60
```

```
Qy 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGAATGGCAAGGAGCGCT 120
Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGAATGGCAAGGAGCGCT 120
```

```
Qy 121 ACCATGGGCTGGCTGCTAGAGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
Db 121 ACCATGGGCTGGCTGCTAGAGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
```

```
Qy 181 GATTCCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGC 240
Db 181 GATTCCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGC 240
```

```
Qy 241 TGGAAAGGATGCAAGGTATGAGTACCTCTGCAATGTGATGACTGTGGATGGCTCCCAAGA 300
Db 241 TGGAAAGGATGCAAGGTATGAGTACCTCTGCAATGTGATGACTGTGGATGGCTCCCAAGA 300
```

```
Qy 301 GATTCAGAGGAGCACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAGAGGAGCACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
```

```
Qy 361 GCTAATATGTTTCAAGCAAGGACTGAGCTAGGATTTATGCAAGATGTTGGAATAAA 420
Db 361 GCTAATATGTTTCAAGCAAGGACTGAGCTAGGATTTATGCAAGATGTTGGAATAAA 420
```

421 ACCTGGCAGGCTTCCCTGGAGTTTGGATACACGATGATGCCCGACACCTTGGCT 480
421 ACCTGGCAGGCTTCCCTGGAGTTTGGATACACGATGATGCCCGACACCTTGGCT 480
481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
541 CGAGATGTTATTAAGACATGCTTGGCCCTGAATAGCAGTGGCAGAGCAATTTGTAC 600
541 CGAGATGTTATTAAGACATGCTTGGCCCTGAATAGCAGTGGCAGAGCAATTTGTAC 600
601 TCCTGTGAGTGGCTCTTATATGTTGGCCCTTCAAAAGCCCAATATACAGAAATCCGA 660
601 TCCTGTGAGTGGCTCTTATATGTTGGCCCTTCAAAAGCCCAATATACAGAAATCCGA 660
661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATGATGATTCCTGGAAAGATTAAG 720
661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATGATGATTCCTGGAAAGATTAAG 720
721 AGTATCTTGGATGGACATCTTTAAACAGGAGAGAAATTTGATGTTGCTGACAGGG 780
721 AGTATCTTGGATGGACATCTTTAAACAGGAGAGAAATTTGATGTTGCTGACAGGG 780
781 GGTGGAATGACCCAGATATGTTAGTATGTCGCACTTGGCCCTCAGCTGGAAATCAGAA 840
781 GGTGGAATGACCCAGATATGTTAGTATGTCGCACTTGGCCCTCAGCTGGAAATCAGAA 840
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATCATGCTTAATGACCTC 900
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATCATGCTTAATGACCTC 900
901 CGACATCAGGCTCAAGCCAAAGCTCTCTTCCAGGATAGGAGTAAATTCGCATCAAT 960
901 CGACATCAGGCTCAAGCCAAAGCTCTCTTCCAGGATAGGAGTAAATTCGCATCAAT 960
961 CAGGACCCCTTGGGCAAGCAAGGTACACAGCTTAGACAGGAGACAACTTGAAGTGTG 1020
961 CAGGACCCCTTGGGCAAGCAAGGTACACAGCTTAGACAGGAGACAACTTGAAGTGTG 1020
1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCAGGAGAGATGGT 1080
1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCAGGAGAGATGGT 1080
1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTATCT 1140
1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTATCT 1140
1141 GCCTGCTTCATACACAGCTCTCTCCCTGTGAAAGAGGAGTGGCTGTATGAGT 1200
1141 GCCTGCTTCATACACAGCTCTCTCCCTGTGAAAGAGGAGTGGCTGTATGAGT 1200
1201 TCAAGGTTAAGAGTACATAAATCCACAGGACCTGTTTGTCTTCAAGTACAGAAACACA 1260
1201 TCAAGGTTAAGAGTACATAAATCCACAGGACCTGTTTGTCTTCAAGTACAGAAACACA 1260
1261 ATGT 1264
1261 ATGT 1264

RESULT 9
US-10-602-219-13
; Sequence 13, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B

; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-13

Query Match 99.8%; Score 1264; DB 7; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGACGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTTCGCTTCTCGGCC 60
Db 1 ATGACGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTTCGCTTCTCGGCC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCCTGGGCAATGGATGGCAAGGACGCT 120
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCCTGGGCAATGGATGGCAAGGACGCT 120
QY 121 ACCATGGGCTGCTGCTGCTGAGGAGGCTTCAATGCAACCTTGAAGTGGCAAGGAGCA 180
Db 121 ACCATGGGCTGCTGCTGCTGAGGAGGCTTCAATGCAACCTTGAAGTGGCAAGGAGCA 180
QY 181 GATTCCTGCTCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAAGC 240
Db 181 GATTCCTGCTCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAAGC 240
QY 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGATGGCTCCCAAGA 300
Db 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGATGGCTCCCAAGA 300
QY 301 GATTCAGAGGAGCAGACTTTCAGGACACCTTCAGGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAGAGGAGCAGACTTTCAGGACACCTTCAGGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATATTGTTACAGCAAAAGGACTGAAGCTAGGAGTTATGCAAGATGTTGGAATAAA 420
Db 361 GCTAATATTGTTACAGCAAAAGGACTGAAGCTAGGAGTTATGCAAGATGTTGGAATAAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACAGCAATGATGCTCCAGACCTTGTCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACAGCAATGATGCTCCAGACCTTGTCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGCAGTGGCAGAGCAATTTGTAC 600

QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGACACAGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGACACAGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATTAATGACCTC 900
DB 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATTAATGACCTC 900
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGAGCTAAATTCGCCATCAAT 960
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGAGCTAAATTCGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGATACAGCTTTAGACAGGAGACAACTTTGAAAGTGTG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGATACAGCTTTAGACAGGAGACAACTTTGAAAGTGTG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCAGGAGATTTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCAGGAGATTTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCCCTGTAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCCCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCGCTCAGCTAGAAAACACA 1260
DB 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCGCTCAGCTAGAAAACACA 1260
QY 1261 ATGT 1264
DB 1261 ATGT 1264

RESULT 11
US-10-851-388-13
; Sequence 13, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-851-388-13
Query Match 99.8%; Score 1264; DB 8; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60

DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTTGGCAAGGACGCT 120
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTTGGCAAGGACGCT 120
QY 121 ACCATGGGCTGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCTGCAGGAAGAGCCA 180
DB 121 ACCATGGGCTGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCTGCAGGAAGAGCCA 180
QY 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGGAAAGATGCAAGGTTATGAGTACCTCTGCAATGATGACTGTTGATGGCTCCCAAGA 300
DB 241 TGGAAAGATGCAAGGTTATGAGTACCTCTGCAATGATGACTGTTGATGGCTCCCAAGA 300
QY 301 GATTCAGAAAGCAGACTTCAGGAGACCTCAGGCGCTTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAGAAAGCAGACTTCAGGAGACCTCAGGCGCTTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTATGTTACAGCAAAAGGACTGAAGCTAGGGATTTATGCAAGATGTTGGAAATAAA 420
DB 361 GCTAATTATGTTACAGCAAAAGGACTGAAGCTAGGGATTTATGCAAGATGTTGGAAATAAA 420
QY 421 ACCTGGCAGAGCTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGAGCTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTTCATGTTGTTACTGTGACAGTTTGGAAATTTG 540
DB 481 GACTGGGAGTAGATCTGCTAAATTTTCATGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 541 GCAGATGTTTAAAGCACATGTCCTTGGCCCTGAAATAGGACTGGCAGAAAGCAATTTGTAC 600
DB 541 GCAGATGTTTAAAGCACATGTCCTTGGCCCTGAAATAGGACTGGCAGAAAGCAATTTGTAC 600
QY 601 TCCTGTGATGGCCCTCTTTTATATGTCGCTTTCAGAAAGCCCAATATACAGAAATCCGA 660
DB 601 TCCTGTGATGGCCCTCTTTTATATGTCGCTTTCAGAAAGCCCAATATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTTCTGGAAGATATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTTCTGGAAGATATAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATGTTGATGTTGCTGACACAGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATGTTGATGTTGCTGACACAGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATTAATGACCTC 900
DB 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATTAATGACCTC 900
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTTCAGGATAAGAGCTAAATTCGCCATCAAT 960
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTTCAGGATAAGAGCTAAATTCGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGATACAGCTTTAGACAGGAGACAACTTTGAAAGTGTG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGATACAGCTTTAGACAGGAGACAACTTTGAAAGTGTG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCAGGAGATTTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCAGGAGATTTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCCCTGTAATCCT 1140

Db 1081 GGACCTCGCTTTATACCATGCGAGTGTGTTCCCTGGGTAAGGAGTGGCCTGTATCTCT 1140
Qy 1141 GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAGAGAGCTAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAGAGAGCTAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAACACA 1260
Db 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGT 1264
Db 1261 ATGT 1264

RESULT 12
US-10-984-389-13
; Sequence 13, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-984-389-13

Query Match 99.8%; Score 1264; DB 9; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGAGCTGAGGAACCCAGAACTACATCTGGCTCGCGCTTGGCTTGGCTTCTCTGCCC 60
Db 1 ATGAGCTGAGGAACCCAGAACTACATCTGGCTCGCGCTTGGCTTGGCTTCTCTGCCC 60
Qy 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 121 ACCATGGCTGGCTGCACTGGAGAGCTTTCATGTCGCACTTGTGCTGCGAGGAGGCA 180
Db 121 ACCATGGCTGGCTGCACTGGAGAGCTTTCATGTCGCACTTGTGCTGCGAGGAGGCA 180
Qy 181 GATTCCTGCATCAGTGAGAGCTTTCATGGAGAGGCGAGAGCTCATGCTCTCAGAAGGC 240
Db 181 GATTCCTGCATCAGTGAGAGCTTTCATGGAGAGGCGAGAGCTCATGCTCTCAGAAGGC 240
Qy 241 TGGAGAGGATGAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGGAGAGGATGAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTCAGAGGAGCACTTCAGGAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAGAGGAGCACTTCAGGAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTCAGCAAGAGGCTGAGCTAGGAGGATTTATGCAAGATGTTGGAATAAA 420
Db 361 GCTAATTTATGTCAGCAAGAGGCTGAGCTAGGAGGATTTATGCAAGATGTTGGAATAAA 420

Qy 421 ACCTGCGAGGCTTCCCTGGGAGTGTGATAGCATTTGATGATGATGATGATGATGATGATGAT 480
Db 421 ACCTGCGAGGCTTCCCTGGGAGTGTGATAGCATTTGATGATGATGATGATGATGATGATGAT 480
Qy 481 GACTGGGAGTAGATCTGCTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATG 540
Qy 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGATAGGACTGGCAGAGCAATTTGATGATG 600
Db 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGATAGGACTGGCAGAGCAATTTGATGATG 600
Qy 601 TCCTGTGAGTGGCTTCTTTATATGAGGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 601 TCCTGTGAGTGGCTTCTTTATATGAGGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Qy 661 CAGTACTGCAATCACTGCGGAAATTTTCTGACATTTGATGATGATGATGATGATGATGATGATG 720
Db 661 CAGTACTGCAATCACTGCGGAAATTTTCTGACATTTGATGATGATGATGATGATGATGATGATG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATGATGATGATGATGATG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATGATGATGATGATGATG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTGTGATGATGATGATGATGATGATGATGATGATGATG 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTGTGATGATGATGATGATGATGATGATGATGATGATG 840
Qy 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Qy 901 CGACATCATGCGCTTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGCTAATTTGCCATCAAT 960
Db 901 CGACATCATGCGCTTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGCTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGGAGTACAGCTTAGACAGGAGAGCACTTTGAGAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGGAGTACAGCTTAGACAGGAGAGCACTTTGAGAGTGTGG 1020
Qy 1021 GAAAGCCTCTCTCAGGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 1021 GAAAGCCTCTCTCAGGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Qy 1081 GGACCTCCTCTTATACCATCGAGTGTCTTCCCTGGGTAAGGAGTGGCTGTAATCTCT 1140
Db 1081 GGACCTCCTCTTATACCATCGAGTGTCTTCCCTGGGTAAGGAGTGGCTGTAATCTCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAGAGCTAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAGAGCTAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAACACA 1260
Db 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGT 1264
Db 1261 ATGT 1264

RESULT 13
US-09-993-059-7
; Sequence 7, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059

```
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(1275)
US-09-993-059-7

Query Match          99.8%; Score 1263.4; DB 3; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTTCCTGGCC 60

QY 61 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACATGGATTTGGCAAGGCGCT 120
DB 61 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACATGGATTTGGCAAGGCGCT 120

QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCCA 180

QY 181 GATTCCTGCATCAGTGAGAACTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCCTGCATCAGTGAGAACTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
DB 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300

QY 301 GATTCAGAGGCGAGACTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAGAGGCGAGACTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 361 GCTAATTATGTTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAAGATGTTTGGAAATAA 420
DB 361 GCTAATTATGTTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAAGATGTTTGGAAATAA 420

QY 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATAGTACGATGATGATGATGATGATGATGATGAT 480
DB 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATAGTACGATGATGATGATGATGATGATGATGAT 480

QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540

QY 541 GCAGATGGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC 600
DB 541 GCAGATGGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC 600

QY 601 TCTGTAGTGGCTCTTTATATGATGGCCCTTTCAAAGGCCCAATATATACAGAAATCCGA 660
DB 601 TCTGTAGTGGCTCTTTATATGATGGCCCTTTCAAAGGCCCAATATATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCAGCTGGGAAATTTGCTGACATTTGATGATGATGATGATGATGATGATG 720
DB 661 CAGTACTGCAATCAGCTGGGAAATTTGCTGACATTTGATGATGATGATGATGATGATGATG 720

QY 721 AGTATCTTGGAGTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
DB 721 AGTATCTTGGAGTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780

QY 781 GGTGGATGACCCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
DB 781 GGTGGATGACCCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840

QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTATGCTTAATGACCTC 900
```

```
DB 841 GFAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTTCATGCTTAATGACCTC 900
QY 901 CGACATCATGCGCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGTAAATTCGCCATCAAT 960
DB 901 CGACATCATGCGCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGTAAATTCGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAAACACCTCTCTCAGGCTTAGCCCTTAGCTGTAGCTATGATATAAAACCGCAGGAGATTGGT 1080
DB 1021 GAAACACCTCTCTCAGGCTTAGCCCTTAGCTGTAGCTATGATATAAAACCGCAGGAGATTGGT 1080
QY 1081 GGACCTCTCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGTAGTGGCTGTAAATCCT 1140
DB 1081 GGACCTCTCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGTAGTGGCTGTAAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAGTGGTTCATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAGTGGTTCATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAAGTACATAAATCCCAAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
DB 1201 TCAAGGTTAAGAAGTACATAAATCCCAAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
QY 1261 ATGTA 1265
DB 1261 ATGCA 1265

RESULT 14
US-10-103-327-7
; Sequence 7, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Tomo H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(1275)
US-10-103-327-7

Query Match          99.8%; Score 1263.4; DB 5; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTTCCTGGCC 60

QY 61 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACATGGATTTGGCAAGGCGCT 120
DB 61 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACATGGATTTGGCAAGGCGCT 120

QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCCA 180
```

QY 181 GATTCTGATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||
QY 181 GATTCTGATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||
QY 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATGATGACTGTTGATGGCTCCCAAGA 300
Db |||||
QY 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATGATGACTGTTGATGGCTCCCAAGA 300
Db |||||
QY 301 GATTGAGAGGAGAGCTTCAGGAGACCTCTCAGCGCTTTCCTCATGGGATTCGCAGCTA 360
Db |||||
QY 301 GATTGAGAGGAGAGCTTCAGGAGACCTCTCAGCGCTTTCCTCATGGGATTCGCAGCTA 360
Db |||||
QY 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
Db |||||
QY 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
Db |||||
QY 421 ACCTGGCAGAGCTTCCCTGGGAGTTTGGATACAGCAATGATGCGCCAGACCTTTCCT 480
Db |||||
QY 421 ACCTGGCAGAGCTTCCCTGGGAGTTTGGATACAGCAATGATGCGCCAGACCTTTCCT 480
Db |||||
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTTG 540
Db |||||
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTTG 540
Db |||||
QY 541 GCAGATGTTTAAAGCACAATGCTTGGCCCTGGAATAGGACTGGCAGAAATTTGTGAC 600
Db |||||
QY 541 GCAGATGTTTAAAGCACAATGCTTGGCCCTGGAATAGGACTGGCAGAAATTTGTGAC 600
Db |||||
QY 601 TCCTGTGATGGCTCTTTATATGTTGGCCCTTCAAAGCCCAATATACAGAAATCCGA 660
Db |||||
QY 601 TCCTGTGATGGCTCTTTATATGTTGGCCCTTCAAAGCCCAATATACAGAAATCCGA 660
Db |||||
QY 661 CAGTACTGCAATCACTGCGAAATTTGCTGACATTTGATGATTCCTGGAAGATATAAG 720
Db |||||
QY 661 CAGTACTGCAATCACTGCGAAATTTGCTGACATTTGATGATTCCTGGAAGATATAAG 720
Db |||||
QY 721 AGTATCTTGGACTGCACTCTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db |||||
QY 721 AGTATCTTGGACTGCACTCTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db |||||
QY 781 GGTGGATGACCAAGATATGTAGTATGTCATCTTGGCACTTGGCTCAGCTGGAATCAGAA 840
Db |||||
QY 781 GGTGGATGACCAAGATATGTAGTATGTCATCTTGGCACTTGGCTCAGCTGGAATCAGAA 840
Db |||||
QY 841 GTAACCTAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
Db |||||
QY 841 GTAACCTAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
Db |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGTAAATGGCCATCAAT 960
Db |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGTAAATGGCCATCAAT 960
Db |||||
QY 961 CAGGACCTTGGGCAAGAGGATACAGCTTAGACAGGGAGACAACTTTGAAGTGG 1020
Db |||||
QY 961 CAGGACCTTGGGCAAGAGGATACAGCTTAGACAGGGAGACAACTTTGAAGTGG 1020
Db |||||
QY 1021 GAACGACCTCTCTCAGGCTTACCTGGCTGTAGCTATGATAAACCAGGAGAGTTGGT 1080
Db |||||
QY 1021 GAACGACCTCTCTCAGGCTTACCTGGCTGTAGCTATGATAAACCAGGAGAGTTGGT 1080
Db |||||
QY 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Db |||||
QY 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Db |||||
QY 1141 GCCTGCTTCATCAGCAGCTCTCTGTGAAAGAGAGCTAGGGTCTATGAAATGACT 1200
Db |||||
QY 1141 GCCTGCTTCATCAGCAGCTCTCTGTGAAAGAGAGCTAGGGTCTATGAAATGACT 1200
Db |||||
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTCTTCAGCTAGAAACACA 1260
Db |||||
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTCTTCAGCTAGAAACACA 1260
Db |||||

QY 1261 ATGTA 1265
Db |||||
QY 1261 ATGCA 1265
Db |||||

RESULT 15

US-10-602-219-7
; Sequence 7, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LBSC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-7

Query Match 99.8%; Score 1263.4; DB 7; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
Db |||||
QY 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
Db |||||
QY 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Db |||||
QY 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Db |||||
QY 121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGAATGCTCAGAGAGCA 180
Db |||||
QY 121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGAATGCTCAGAGAGCA 180
Db |||||
QY 181 GATTCTGCACTAGTGAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||
QY 181 GATTCTGCACTAGTGAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||
QY 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATGATGACTTGGATGGCTCCCAAGA 300
Db |||||
QY 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATGATGACTTGGATGGCTCCCAAGA 300
Db |||||
QY 301 GATTGAGAGGAGCACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db |||||

```
Db      ||||| 301 GATTGAGAGGCAGACTTCAGGAGACCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY      ||||| 361 GCTAATATATGTTACAGCAAGAGACTGAAGCTAGGGATTTATGACAGATGTTGGAATAAAA 420
Db      ||||| 361 GCTAATATATGTTACAGCAAGAGACTGAAGCTAGGGATTTATGACAGATGTTGGAATAAAA 420
QY      ||||| 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACTAGCAGATTCAGGACCTTTGCT 480
Db      ||||| 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACTAGCAGATTCAGGACCTTTGCT 480
QY      ||||| 481 GACTGGGAGTAGACTCTGTAATAATTTGATGGTTGTTACTGACAGATTTGGAATAATTTG 540
Db      ||||| 481 GACTGGGAGTAGACTCTGTAATAATTTGATGGTTGTTACTGACAGATTTGGAATAATTTG 540
QY      ||||| 541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
Db      ||||| 541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
QY      ||||| 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db      ||||| 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY      ||||| 661 CAGTACTGCAATCAGTCGGGCAAAATTTGCTGACATTTGATGATTCCTGGAAAAAGTATAAAG 720
Db      ||||| 661 CAGTACTGCAATCAGTCGGGCAAAATTTGCTGACATTTGATGATTCCTGGAAAAAGTATAAAG 720
QY      ||||| 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGCACAGGG 780
Db      ||||| 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGCACAGGG 780
QY      ||||| 781 GGTGGGAATGACCCAGATATCTTAGTATGGCACTTTGGCCTCAGCTGGGAATCAGCAA 840
Db      ||||| 781 GGTGGGAATGACCCAGATATCTTAGTATGGCACTTTGGCCTCAGCTGGGAATCAGCAA 840
QY      ||||| 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCTAATGACCTC 900
Db      ||||| 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCTAATGACCTC 900
QY      ||||| 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAAATGCCATCAAT 960
Db      ||||| 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAAATGCCATCAAT 960
QY      ||||| 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGCAGGAGACAACTTTGAACTGTGG 1020
Db      ||||| 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGCAGGAGACAACTTTGAACTGTGG 1020
QY      ||||| 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080
Db      ||||| 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080
QY      ||||| 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
Db      ||||| 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
QY      ||||| 1141 GCCTGCTTTCATCACAGCTCTCCTCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200
Db      ||||| 1141 GCCTGCTTTCATCACAGCTCTCCTCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200
QY      ||||| 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
Db      ||||| 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
QY      ||||| 1261 ATGTA 1265
Db      ||||| 1261 ATGCA 1265
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 08:30:05 ; Search time 284 Seconds
(without alignments)
2313.369 Million cell updates/sec

Title: US-10-602-219-11
Perfect score: 1266
Sequence: 1 atgcagctggaaccacaga.....agctagaacacacatatga 1266

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4168288 seqs, 259477437 residues

Total number of hits satisfying chosen parameters: 8336576

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA_New:*
1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37	2.9	2779	6	US-10-750-185-40681
C 2	36.6	2.9	1306	6	US-10-750-185-42255
C 3	35	2.8	321019	6	US-10-995-561-13204
C 4	34	2.7	1932	6	US-10-750-185-55363
C 5	33	2.6	1437	6	US-10-750-185-43877
C 6	32.8	2.6	1666	6	US-10-821-234-149
C 7	32.8	2.6	146656	7	US-11-121-086-68
C 8	32.8	2.6	188682	7	US-11-112-908-23
C 9	32.8	2.6	190882	7	US-11-121-086-69
C 10	32.4	2.6	1685	6	US-10-750-185-53897
C 11	32.2	2.5	86081	6	US-10-995-561-13246
C 12	32	2.5	1254	6	US-10-750-185-29504
C 13	31.8	2.5	201	6	US-10-995-561-29100
C 14	31.8	2.5	4847	6	US-10-750-185-60280
C 15	31.6	2.5	2787	6	US-10-841-129-3
C 16	31.6	2.5	1125000	6	US-10-995-561-13286
C 17	31.4	2.5	119160	7	US-11-121-086-12
C 18	31.4	2.5	161874	7	US-11-121-086-75
C 19	31.2	2.5	1522	6	US-10-750-185-37196
C 20	31	2.4	854	6	US-10-750-185-28311
C 21	31	2.4	1084	6	US-10-750-185-61883
C 22	31	2.4	1783	6	US-10-750-185-37071
C 23	31	2.4	168516	7	US-11-121-086-3

C 24	30.8	2.4	944	6	US-10-750-185-64685	Sequence 64685, A
C 25	30.8	2.4	1806	6	US-10-131-826A-121	Sequence 121, App
C 26	30.8	2.4	5234	6	US-10-821-234-357	Sequence 357, App
C 27	30.6	2.4	1131	6	US-10-750-185-54302	Sequence 54302, A
C 28	30.6	2.4	40394	6	US-10-995-561-13493	Sequence 13493, A
C 29	30.4	2.4	2103	6	US-10-750-185-42925	Sequence 42925, A
C 30	30.4	2.4	163162	7	US-11-121-086-66	Sequence 66, Appl
C 31	30.4	2.4	191091	7	US-11-121-086-60	Sequence 60, Appl
C 32	30.2	2.4	1805	6	US-10-750-185-39400	Sequence 39400, A
C 33	30.2	2.4	126552	7	US-11-121-086-1	Sequence 1, Appl
C 34	30.2	2.4	317876	6	US-10-995-561-13227	Sequence 13227, A
C 35	30	2.4	1267	6	US-10-750-185-43163	Sequence 43163, A
C 36	29.8	2.4	858	6	US-10-750-185-28504	Sequence 28504, A
C 37	29.8	2.4	2242	6	US-10-750-185-39271	Sequence 39271, A
C 38	29.8	2.4	3021	6	US-10-750-185-35062	Sequence 35062, A
C 39	29.8	2.4	3821	6	US-10-750-185-63028	Sequence 63028, A
C 40	29.8	2.4	7980	6	US-10-509-921-4	Sequence 4, Appl
C 41	29.8	2.4	7980	6	US-10-509-921-5	Sequence 5, Appl
C 42	29.8	2.4	7989	6	US-10-509-921-2	Sequence 2, Appl
C 43	29.8	2.4	7989	6	US-10-509-921-6	Sequence 6, Appl
C 44	29.8	2.4	7989	6	US-10-509-921-13	Sequence 13, Appl
C 45	29.8	2.4	7989	6	US-10-509-921-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-10-750-185-40681/c
; Sequence 40681, Application US/10750185
; Publication No. US200502603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10750185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40681
; LENGTH: 2779
; TYPE: DNA
; ORGANISM: Bovine 19866880781246
US-10-750-185-40681

Query Match	2.9%	Score 37;	DB 6;	Length 2779;
Best Local Similarity	52.2%	Pred. No. 0.082;	Mismatches 75;	Indels 0;
Matches	82;	Conservative 0;		
Qy	631	TTTCAAAGCCAAATATACAGAAATCCGACAGTACTGCAATCACTGGCGAAATTTTGT	690	
Db	2693	TTTTAGACTCCCTGTTTGATTTGTTTCTTATTTCTTATTTATGCGAGTTCTACTGTTT	2634	
Qy	691	GACATTGATGATTCCTGGAAAAGTATAAAGAGTATCTTTGGACTGGACATCTTTTAAACGAG	750	
Db	2633	GACTCTTGGGCTCAGGATACACTGTAACCTTCAGGAGGCTTGTGTGATCTGTATACCAG	2574	
Qy	751	GAGAGATTGTTGATTTGCTGGACCGGGGGTTTGA	787	
Db	2573	CAAGGAGTATTTCATGTTGCTGTACCAAGTTTGA	2537	

RESULT 2
US-10-750-185-42255
; Sequence 42255, Application US/10750185

; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42255
; LENGTH: 1306
; TYPE: DNA
; ORGANISM: Bovine 19866880955804
US-10-750-185-42255

Query Match 2.9%; Score 36.6; DB 6; Length 1306;
Best Local Similarity 53.1%; Pred. No. 0.067;
Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 715 ATAAGAGTATCTTGACGTGGAGCATCTTTTAACGAGGAGAGAAATTTGTCATGTTGCTGA 774
Db ATGCGAGTTAAATAATAGACAACTTTAGTCAAGAGAGCAATTTGTTGTAAGGAATTC 643
QY 775 CCAGGGGTTGGAATGACCCAGCATATGTTAGTGATTGGCAACTTTGGCCTCAGCTGGAAT 834
Db CAACCCCTGGAACTTGAGTGTCTGTAATATAGGCGAGAGCTGGCCTTCCTTGCCTT 703
QY 835 CAGCAAGTAACCTCAGATGCCCTCTGG 861
Db CTGTCCCTGCCTCAGGTGGTCCCTGG 730

RESULT 3
US-10-995-561-13204/c
; Sequence 13204, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13204
; LENGTH: 321019
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(321019)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13204

Query Match 2.8%; Score 35; DB 6; Length 321019;
Best Local Similarity 49.1%; Pred. No. 11;
Matches 86; Conservative 2; Mismatches 87; Indels 0; Gaps 0;
QY 915 TCAGCAAGGCTCTCTTCAGATGAAGGACGTATTTGCCATCATCAGGACCCCTTGGG 974
Db TCACAGAGGCTGACGTAGCAGAACACRAAGATTTGTAAAGACGAGGATGCTTAAG 301318
QY 975 CAAGCAAGGGTACCAGCTTAGACAGGAGGACAACTTTGAAGTGTGGGAACGACCTCTCTC 1034

Db 301317 CATGAGGCTGCCTGCCTTTGTGAGGCAGTGAACATGAGTCCCTTGGGAAGAAACCTCTT 301258
QY 1035 AGGCTTAGCCTGGGCTGTAGCTATCATATAAACCGCAGCAGAGATTGTTGACCTCGC 1089
Db 301257 AGCAGAAGCTCAACACAGCAGCTGTCAAAGATCTGCAGGTGCACCTGGGAGCTGGGC 301203
RESULT 4
US-10-750-185-55363/c
; Sequence 55363, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 55363
; LENGTH: 1932
; TYPE: DNA
; ORGANISM: Bovine 19866880928582
US-10-750-185-55363

Query Match 2.7%; Score 34; DB 6; Length 1932;
Best Local Similarity 57.5%; Pred. No. 0.7;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 655 ATCCGACAGTACTGCAATCACTGGCGAAATTTTGCTGACATTGATGATTCCTGGAAAGT 714
Db ATCTGAGATTTCTGAGATGATTTTCCAAATTTATTTGGAGAAAGGATTCCTGAAGAATT 1022
QY 715 ATAAAGAGTATCTTGGACTGGACATCTTTTAAACGAGGAGAGAATTG 760
Db AGTGAATTTCTTGATGTACATCTTTTATGACATTATAATTG 976

RESULT 5
US-10-750-185-43877/c
; Sequence 43877, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 43877
; LENGTH: 1437
; TYPE: DNA
; ORGANISM: Bovine 19866880874544
US-10-750-185-43877

```
Query Match      2.6%; Score 33; DB 6; Length 1437;
Best Local Similarity 53.5%; Pred. No. 1.3;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 835 CAGCAAGTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAAT 894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 CAGGAATTTATTTGGTGTGGCTGACTGGGATATCCAGGAGGAGCTGGCTCAGGAATTC 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 895 GACCTCCGACACATCAGCCCTCAAGCCCAAGCTCTCTCTTCAGGATAAGGACGTAAATGCC 954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 GAGCCAGAGCCCTTCTGGGCTCCCTCTCATCTCTCTCCCGTGCTAGATGCTCTTTGGA 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 955 ATCAATCAG 963
    ||| ||| |||
Db 338 ATGGAACAG 330
    ||| ||| |||

RESULT 6
US-10-821-234-149
; Sequence 149, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Presclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 149
; LENGTH: 1666
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-149

Query Match      2.6%; Score 32.8; DB 6; Length 1666;
Best Local Similarity 64.5%; Pred. No. 1.6;
Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 156 CAACCTTGACTCCAGGAGAGCCAGATTCTCTGCATCATGAGAGCTCTTCATGGAGAT 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 CAGCTTGCTCTCCCTGGATATCTCCGATTACTCCAAATGTAAAGATCATGATTGAGAG 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 216 GGCAGAGCTCATGGTC 231
    ||| ||| ||| |||
Db 298 AGGAGAGCTTTTCTC 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-11-121-086-68
; Sequence 68, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-0000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 68
; LENGTH: 146656
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-68

Query Match      2.6%; Score 32.8; DB 7; Length 188682;
Best Local Similarity 59.8%; Pred. No. 43;
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 256 TATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGAGATTTCAGAGGCGAGA 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139854 TATGCTATCTCTGGATTATCACTGTAGTTGGGATAAAGGAATCTGACTGCCGAGA 139795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 316 CTTGAGGAGACCTCTCAGCGCTTTCTCTCATCG 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139794 TCTGAGACATAGTCCACCCCTTTTGCACTGGG 139763
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-11-121-086-69
; Sequence 69, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-0000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 69
; LENGTH: 190882
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-69

Query Match      2.6%; Score 32.8; DB 7; Length 146656;
Best Local Similarity 58.0%; Pred. No. 37;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 644 ATTATACAGAATCCGACAGTACTGCAATCACTCTCTCGGAATTTTGCTGACATTGATGATT 703
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134039 ATTACAGCATATGCCAAAGTACTAGGCTCACTTGTCTAATTTTAGACCCCAATGATGATT 134098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 704 CCTGGAAAGTATAAAGAGATATCTTTGGACTGGACATCTTT 743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134059 CTTCTTATATTAATCACTTATCTCTTAATAGAAATTTATT 134138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-11-112-908-23/c
; Sequence 23, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 23
; LENGTH: 188682
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-23
```


US-11-121-086-69

Query Match 2.6%; Score 32.8; DB 7; Length 190882;
Best Local Similarity 58.0%; Pred. No. 44; Mismatches 42; Indels 0; Gaps 0;
Matches 58; Conservative 0;

QY 644 ATTATACAGAAATCCGACAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATT 703
|||||
Db 22276 ATTACAGCATATGCCAAGTACTAGGCTCACTTGTCTAATTTTAGCACCAATGATGATT 22335
|||||

QY 704 CTGGAAAAGTAAAGAGTATCTTGGACTGGACATCTTTT 743
|||||
Db 22336 CTTCTTATTAATCACTTATCTTAAATAGAAATTAAT 22375
|||||

RESULT 10

US-10-750-185-53897
; Sequence 53897, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 53897
; LENGTH: 1685
; TYPE: DNA
; ORGANISM: Bovine 19866880964900
US-10-750-185-53897

Query Match 2.6%; Score 32.4; DB 6; Length 1685;
Best Local Similarity 60.0%; Pred. No. 2.3; Mismatches 36; Indels 0; Gaps 0;
Matches 54; Conservative 0;

QY 127 GCGTCTGCTGCTGAGGCGCTTCATGCTCAACCTTGACTGCCAGGAGCGCAGATTCC 186
|||||
Db 849 GCGTTCTGCTGATGGCTGCTGTTTATGAGCCACGTTTAAAGGAGAGGAAACCAACATTTT 908
|||||

QY 187 TGCATCAGTGAGAAGCTCTTCATGGAGATG 216
|||||
Db 909 TGTTCCTTGAGTAACTCGTTATAGATTGG 938
|||||

RESULT 11

US-10-995-561-13246/c
; Sequence 13246, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13246
; LENGTH: 86081
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13246

Query Match 2.5%; Score 32.2; DB 6; Length 86081;
Best Local Similarity 48.0%; Pred. No. 41; Mismatches 86; Indels 0; Gaps 0;
Matches 82; Conservative 3;

QY 104 GATTGGCAAGAGCGCTTACCATTGGCTGCTGCACTGGGAGCGCTTCAATGTGCAACCTTG 163
|||||
Db 68847 GGTGGGCAAGGCTTTGACCTTGGCCTGGCGTTAGTGGCGTGAGTAGGTACCTG 68788
|||||

QY 164 ACTGCCAGGAAGACGAGATTCTCTGCATCAGTGAGAAGCTCTTTCATGGAGATGGCAGAGC 223
|||||
Db 68787 CGTCTGGGGTYGGACTGAGGCGCAGCATCAGCATGTTGTTCTCCAGGCGGCTGCARGTG 68728
|||||

QY 224 TCATGCTCTCAGAAGCGCTGGAAGGATGCAGGTTATGAGTACCTCTGCAATTG 274
|||||
Db 68727 GCAGGCTGCCATCCAGCTGYAAATGCACTAGCAGGGCCCTGGCCCTG 68677
|||||

RESULT 12

US-10-750-185-29504
; Sequence 29504, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 29504
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Bovine 19866880446821
US-10-750-185-29504

Query Match 2.5%; Score 32; DB 6; Length 1254;
Best Local Similarity 46.4%; Pred. No. 2.5; Mismatches 120; Indels 0; Gaps 0;
Matches 104; Conservative 0;

QY 215 TGGCAGAGCTCATGTGCTCAGAAAGCTGGAAGGATGCGAGGTTATGAGTACCTCTGCAATTG 274
|||||
Db 440 TGGAGATCCCATGGACAGAGGAGCGCTGGCAGGCTACAGTCCATGGTGGTAAAGAAATTG 499
|||||

QY 275 ATGACTGTTGGATGGCTCCCAAGAGATTCAAGAGGCGAGCTTCAGGCGAGACCTTCAGC 334
|||||
Db 500 GACACAATTTAGTGAATAACAATGATGCAGAAAAGCCATAATACATTTTGGAGAGAGT 559
|||||

QY 335 GCTTTCTCATGGGATTCGCCAGCTAGCTTAATTATGTTTCACAGCAAGAGGACTGAGAGTAG 394
|||||
Db 560 GGTGAGAGAGCGGAGCCGATAAATGTTTAAATAATATTCATGCGCAACATCTCCGAGCAAA 619
|||||

QY 395 GGATTTATGCGCATGTTGGAATAAAACCTGCGCAGGCTTCCT 438
|||||
Db 620 TTTAAATAATAATAAAAGTAAATAGACTCATTTCTCTGCTACTCT 663
|||||

RESULT 13

US-10-995-561-29100/c
; Sequence 29100, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 08:33:27 ; Search time 187 Seconds
(without alignments)
989.190 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQLNPEHLGICALALRFLA.....RLRSHINPTGTVLLQLENTM 421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2314	100.0	421	5	Aae28210 Human rGA
2	2314	100.0	421	7	ADD84751 Human alp
3	2314	100.0	421	8	ADJ88277 Human WT
4	2314	100.0	421	8	ADM48581 Human wil
5	2314	100.0	421	8	ADU66916 Human alp
6	2314	100.0	421	9	AEA27445 Human alp
7	2314	100.0	425	9	AEA27472 Human alp
8	2314	100.0	427	5	Aae28211 Human rGA
9	2314	100.0	427	5	ADD84753 Human alp
10	2314	100.0	427	8	ADJ88279 Human WT
11	2314	100.0	427	8	ADM48583 Human wil
12	2314	100.0	427	8	ADU66918 Human alp
13	2314	100.0	427	9	AEA27447 Human alp
14	2314	100.0	429	2	AAR53765 Alpha-gal
15	2314	100.0	429	2	AAR63234 Human alp
16	2314	100.0	429	2	AAR70207 Alpha-gal
17	2314	100.0	429	5	Aae28206 Human WT
18	2314	100.0	429	7	ADF76504 Novel hum
19	2314	100.0	429	8	ADM49740 Human alp
20	2314	100.0	429	8	ADP24703 PRO poly
21	2314	100.0	429	8	ADU18066 Human can
22	2314	100.0	429	8	ADU74416 Human alp
23	2314	100.0	429	9	ADY19459 PRO poly
24	2314	100.0	429	9	ADY17095 PRO poly

25	2314	100.0	429	9	AEA27471 Human alp
26	2314	100.0	429	9	AEb43188 Human alp
27	2314	100.0	430	7	ABM79001 Human alp
28	2314	100.0	431	5	Aae28209 Human rGA
29	2314	100.0	431	7	ADD84749 Human alp
30	2314	100.0	431	8	ADJ88275 Human WT
31	2314	100.0	431	8	ADM48679 Human wil
32	2314	100.0	431	8	ADU66914 Human alp
33	2314	100.0	431	9	AEA27443 Human alp
34	2314	100.0	435	5	Aae28207 Human WT
35	2314	100.0	435	7	ADD84745 Human alp
36	2314	100.0	435	8	ADJ88271 Human WT
37	2314	100.0	435	8	ADM48675 Human wil
38	2314	100.0	435	8	ADU66910 Human alp
39	2314	100.0	435	9	AEA27439 Human alp
40	2314	100.0	478	2	RAY48575 Human bre
41	2311	99.9	429	2	AAR07305 Precursor
42	2309	99.8	424	5	Aae28208 Human rGA
43	2309	99.8	424	7	ADD84747 Human alp
44	2309	99.8	424	8	ADJ88273 Human WT
45	2309	99.8	424	8	ADM48677 Human wil

ALIGNMENTS

RESULT 1
AAE28210
ID AAE28210 standard; protein; 421 AA.
XX
AC AAE28210;
XX
DT 27-DEC-2002 (first entry)
XX
DE Human rGAL-8 protein.
XX
KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-8.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Cleavage-site 401..402
FT /note= "CTPP cleavage site"
XX
PN US2002088024-A1.
XX
PD 04-JUL-2002.
XX
PF 13-NOV-2001; 2001US-00993059.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2002-681656/73.
DR N-PSDB; AAD45223.
XX
PT Novel human alpha-galactosidase polypeptide useful for treating lysosomal storage diseases.
XX
PS Claim 7; Page 44; 88pp; English.
XX
CC The invention relates to human alpha-galactosidase truncated at the carboxy terminus and the production of enzymatically active recombinant human and animal lysosomal enzymes. The invention is useful for producing lysosomal enzymes for treating lysosomal storage diseases, producing altered or mutated proteins, enzymatically active or otherwise, to serve as precursors or substrates for further in vivo or in vitro processing to

CC a specialised industrial form for research or therapeutic uses, to
CC produce more effective therapeutic enzyme, for producing antibodies
CC against lysosomal enzymes for medical diagnostic use, and in any
CC commercial process that involves substrate hydrolysis. The present
CC sequence is human rGAL-8 protein
XX
SQ Sequence 421 AA;

Query Match 100.0%; Score 2314; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPELHLCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNCDCQEP 60
DB 1 MOLRNPELHLCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNCDCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGVLLKFDGICYCDLENL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGVLLKFDGICYCDLENL 180
QY 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKPNYTEIROYCNHWRNFADIDDSWKSITK 240
DB 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKPNYTEIROYCNHWRNFADIDDSWKSITK 240
QY 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWRPPLSGLAWAVAMINRQETG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWRPPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 2
ADD84751
ID ADD84751 standard; protein; 421 AA.
XX
AC ADD84751;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human alpha-galactosidase rGAL-8 polypeptide.
XX
KW Human; alpha-galactosidase; rGAL-4; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease; enzyme.
XX
OS Homo sapiens.
XX
PN US2003106095-A1.
XX
PD 05-JUN-2003.
XX
PF 20-MAR-2002; 2002US-00103327.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PR 13-NOV-2001; 2001US-00993059.
XX
XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX

PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2003-801257/75.
DR N-PSDB; ADD84750.
XX
PT New polynucleotide for producing active recombinant human and animal
PT lysosomal enzymes in a plant expression system that can be used in enzyme
PT replacement therapy.
XX
PS Claim 7; SEQ ID NO 12; 77pp; English.
XX
CC The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents a human alpha-galactosidase derivative
CC polypeptide of the invention.
XX
SQ Sequence 421 AA;

Query Match 100.0%; Score 2314; DB 7; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPELHLCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNCDCQEP 60
DB 1 MOLRNPELHLCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNCDCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGVLLKFDGICYCDLENL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGVLLKFDGICYCDLENL 180
QY 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKPNYTEIROYCNHWRNFADIDDSWKSITK 240
DB 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKPNYTEIROYCNHWRNFADIDDSWKSITK 240
QY 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWRPPLSGLAWAVAMINRQETG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWRPPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 3
ADJ88277
ID ADJ88277 standard; protein; 421 AA.
XX
AC ADJ88277;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human WT rGAL-8 (galactosidase).
XX
KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease; human; enzyme.
XX
OS Homo sapiens.
XX

```

PN US2004016021-A1.
XX
PD 22-JAN-2004.
XX
PF 23-JUN-2003; 2003US-00602219.
XX
PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
FI Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
XX
DR N-PSDB; ADJ88276.
XX
PT New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
PS Claim 7; SEQ ID NO 12; 71pp; English.
XX
CC The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC protein.
XX
XX Sequence 421 AA;
XX
Query Match 100.0%; Score 2314; DB 8; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPDLHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFHFMCLDCQEEP 60
DB 1 MOLRNPDLHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFHFMCLDCQEEP 60
QY 61 DSCISEKLFMEAEIWMVSGWKDAGVEYICDDCWAPQDSEGRLOADPQRFPHGIQOL 120
DB 61 DSCISEKLFMEAEIWMVSGWKDAGVEYICDDCWAPQDSEGRLOADPQRFPHGIQOL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDYDIDAQTFADWGVLLKPDGCGYCDLSNLL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDYDIDAQTFADWGVLLKPDGCGYCDLSNLL 180
QY 181 ADGYKXMSIALNRTGRSIVYSCWELPMYPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGYKXMSIALNRTGRSIVYSCWELPMYPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSFNOERIVDVAGPGWNDPDLVIGNFGLSNQOQVQTMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNOERIVDVAGPGWNDPDLVIGNFGLSNQOQVQTMALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQEI 360

```

```
CC invention.
XX
SQ Sequence 421 AA;
Query Match 100.0%; Score 2314; DB 8; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCBE 60
DB 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCBE 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPODSEGRLOADPQPFPHGI 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPODSEGRLOADPQPFPHGI 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGICYD 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGICYD 180
QY 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROYCNHWRNFADIDDS 240
DB 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROYCNHWRNFADIDDS 240
QY 241 SILDWTFSNQRIVDVAGPGGWNDDMLVIGNFGLSWNQVDTOMALWAIMAAPLF 300
DB 241 SILDWTFSNQRIVDVAGPGGWNDDMLVIGNFGLSWNQVDTOMALWAIMAAPLF 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMIN 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMIN 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGT 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGT 420
QY 421 M 421
DB 421 M 421
RESULT 5
ADU66916
ID ADU66916 standard; protein; 421 AA.
XX
AC ADU66916;
XX
DT 10-FEB-2005 (first entry)
DE Human alpha-galactosidase protein #5.
XX
KW Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human.
XX
OS Homo sapiens.
XX
PN US2004234516-A1.
XX
PD 25-NOV-2004.
XX
PF 21-MAY-2004; 2004US-00851389.
XX
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
PR 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
PA
PI Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2004-821274/81.
DR
XX N-PSDB; ADU66915.
PT A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.
XX
PS Disclosure; SEQ ID NO 12; 88pp; English.
XX
CC The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein.
XX
SQ Sequence 421 AA;
Query Match 100.0%; Score 2314; DB 8; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCBE 60
DB 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCBE 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPODSEGRLOADPQPFPHGI 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPODSEGRLOADPQPFPHGI 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGICYD 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGICYD 180
QY 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROYCNHWRNFADIDDS 240
DB 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROYCNHWRNFADIDDS 240
QY 241 SILDWTFSNQRIVDVAGPGGWNDDMLVIGNFGLSWNQVDTOMALWAIMAAPLF 300
DB 241 SILDWTFSNQRIVDVAGPGGWNDDMLVIGNFGLSWNQVDTOMALWAIMAAPLF 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMIN 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMIN 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGT 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGT 420
QY 421 M 421
DB 421 M 421
RESULT 6
AEA27445
ID AEA27445 standard; protein; 421 AA.
XX
AC AEA27445;
XX
DT 11-AUG-2005 (first entry)
XX
DE Human alpha-galactosidase protein, rGAL-8, SEQ ID NO: 12.
XX
KW Gauchers disease; metabolic; neurological disease; niemann pick disease;
KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
KW antilepemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
```


KW	enzyme.	Db	361	GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT	420
XX	Homo sapiens.	Qy	421	M 421	
XX		Db	421	M 421	
PN	US2005125859-A1.				
XX	09-JUN-2005.				
PD					
XX	08-NOV-2004; 2004US-00984389.				
XX	26-JUL-2000; 2000US-00626127.				
PR	13-NOV-2001; 2001US-00993059.				
PR	20-MAR-2002; 2002US-00103327.				
XX	(LARG-) LARGE SCALE BIOLOGY CORP.				
PA	Garger SJ, Turpen TH, Kumagai MH;				
XX	WPI; 2005-404004/41.				
DR	N-PSDB; AEA27444.				
XX	New isolated polypeptides useful for producing lysosomal enzymes in				
PT	plants to be utilized in enzyme replacement therapy or for the				
PT	therapeutic treatment of human or animal lysosomal storage diseases, e.g.				
PT	Gaucher's disease.				
XX					
PS	Claim 28; SEQ ID NO 12; 88pp; English.				
XX					
CC	The present invention relates to the production of human and animal				
CC	lysosomal enzymes in plants by a transient plant expression system. The				
CC	invention relates to glucocerebrosidase (GCB, GCR) and alpha-				
CC	galactosidase (Gal) enzymes having a post-translational modification				
CC	provided by the plant expression system. The invention is useful in				
CC	enzyme replacement therapy for treating lysosomal storage diseases such				
CC	as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs				
CC	disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is				
CC	also useful in researches for developing new approaches to medical				
CC	treatment of lysosomal storage diseases and in industrial processes				
CC	involving enzymatic substrate hydrolysis. The present sequence is the				
CC	human alpha-galactosidase protein. Note: The present sequence is the SEQ				
CC	ID NO: 8 which is shown in page 36-38 of the specification. This sequence				
CC	differs from the SEQ ID NO: 8 given in the sequence listing (see				
XX	AEA27441).				
XX					
XX	Sequence 421 AA;				
Qy	Query Match 100.0%; Score 2314; DB 9; Length 421;				
Db	Best Local Similarity 100.0%; Pred. No. 1.8e-221;				
	Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1 MOLRNPHELGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCNLDQCQEP 60				
Db	1 MOLRNPHELGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCNLDQCQEP 60				
Qy	61 DSCISEKLFMEAEILVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120				
Db	61 DSCISEKLFMEAEILVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120				
Qy	121 ANYVHSGKLGLGIYADVGNKTCAGPPGSGFYGYDDIAQTFADGWDLKFKDGCYCDLSLENL 180				
Db	121 ANYVHSGKLGLGIYADVGNKTCAGPPGSGFYGYDDIAQTFADGWDLKFKDGCYCDLSLENL 180				
Qy	181 ADGQKMSLALNRTGRSIVYCEWPLVYMPFQKPNYTEIRQYCNHWRNFADIDSWKSIK 240				
Db	181 ADGQKMSLALNRTGRSIVYCEWPLVYMPFQKPNYTEIRQYCNHWRNFADIDSWKSIK 240				
Qy	241 SILDWTSTNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300				
Db	241 SILDWTSTNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300				
Qy	301 RHISPOAKALLQDKDVIAINQDPLKQGYQLRQGNFVWERPLSGLAWAVAMINRQIG 360				
Db	301 RHISPOAKALLQDKDVIAINQDPLKQGYQLRQGNFVWERPLSGLAWAVAMINRQIG 360				
Qy	361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420				

```

QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIRQL 120
|
|
|
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIRQL 120
|
|
|
QY 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYGYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180
|
|
|
Db 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYGYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180
|
|
|
QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
|
|
|
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
|
|
|
QY 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
|
|
|
Db 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
|
|
|
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
|
|
|
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
|
|
|
QY 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
|
|
|
Db 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
|
|
|
QY 421 M 421
|
|
Db 421 M 421

RESULT 8
AAE28211
ID AAE28211 standard; protein; 427 AA.
XX
AC AAE28211;
XX
DT 27-DEC-2002 (first entry)
XX
DE Human rGAL-8R protein.
XX
KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-8R.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Cleavage-site 401..402
FT /note= "CTPP cleavage site"
XX
PN US2002088024-A1.
XX
PD 04-JUL-2002.
XX
PF 13-NOV-2001; 2001US-00993059.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUNA/) KUMAGAI M H.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2002-681656/73.
DR N-PSDB; AAD45224.
XX
PT Novel human alpha-galactosidase polypeptide useful for treating lysosomal
PT storage diseases.
XX
PS Claim 7; Page 46-47; 88pp; English.
XX
CC The invention relates to human alpha-galactosidase truncated at the
CC carboxy terminus and the production of enzymatically active recombinant

```

CC human and animal lysosomal enzymes. The invention is useful for producing
 CC lysosomal enzymes for treating lysosomal storage diseases, producing
 CC altered or mutated proteins, enzymatically active or otherwise, to serve
 CC as precursors or substrates for further in vivo or in vitro processing to
 CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-8R protein

XX Sequence 427 AA;

```

Query Match      100.0%; Score 2314; DB 5; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPELHLCALALRFLALVSDIPGARALDNGLARPTTGMWHWRFMCNLDCCQEP 60
|
|
|
Db 1 MOLRNPELHLCALALRFLALVSDIPGARALDNGLARPTTGMWHWRFMCNLDCCQEP 60
|
|
|
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIRQL 120
|
|
|
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIRQL 120
|
|
|
QY 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYGYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180
|
|
|
Db 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYGYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180
|
|
|
QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
|
|
|
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
|
|
|
QY 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
|
|
|
Db 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
|
|
|
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
|
|
|
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
|
|
|
QY 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
|
|
|
Db 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
|
|
|
QY 421 M 421
|
|
Db 421 M 421

```

RESULT 9

ADD84753
 ID ADD84753 standard; protein; 427 AA.

XX AC ADD84753;

XX DT 29-JAN-2004 (first entry)

XX DE Human alpha-galactosidase rGAL-8R polypeptide.

XX KW Human; alpha-galactosidase; rGAL-4; lysosomal enzyme;
 KW enzyme replacement therapy; lysosomal disease; enzyme.

XX OS Homo sapiens.

XX XX US2003106095-A1.

XX XX 05-JUN-2003.

XX XX 20-MAR-2002; 2002US-00103327.

XX XX 26-JUL-2000; 2000US-00626127.

XX XX 13-NOV-2001; 2001US-00993059.


```
|||||
Db 241 SILDWTSTFQRIIVDVAGGGWDPMLVIGNFGLSWQVVTQMALWAIMAAPLPMNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWEPRLSGLAWAVAMINRQIIG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWEPRLSGLAWAVAMINRQIIG 360
QY 361 GPRSYTIAVASIGKGVACNPACFIITQLLPVKRKLGFYEWTSLRSHINPTGTVLQLENT 420
Db 361 GPRSYTIAVASIGKGVACNPACFIITQLLPVKRKLGFYEWTSLRSHINPTGTVLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 11
ADM48683
ID ADM48683 standard; protein; 427 AA.
XX
AC ADM48683;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human wild type rGAL-8R protein.
XX
KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW enzyme.
XX
OS Homo sapiens.
XX
PN US2004023281-A1.
XX
PD 05-FEB-2004.
XX
PF 23-JUN-2003; 2003US-00602220.
XX
PR 26-FEB-1988; 88US-00160766.
XX
PR 26-FEB-1988; 88US-00160771.
XX
PR 17-FEB-1989; 89US-00310881.
XX
PR 22-OCT-1990; 90US-00600244.
XX
PR 31-JUL-1992; 92US-00923692.
XX
PR 30-DEC-1992; 92US-00997733.
XX
PR 29-DEC-1993; 93US-00176414.
XX
PR 19-JAN-1994; 94US-00184237.
XX
PR 14-OCT-1994; 94US-00324003.
XX
PR 21-MAY-1999; 99US-00316572.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PR 13-NOV-2001; 2001US-00993059.
XX
(TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
PA (POGU/) POGUE G P.
PA (ERWT/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-142650/14.
XX
DR N-PSDB; ADM48682.
XX
PT New alpha-galactosidase polypeptides, useful in producing recombinant
XX
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
XX
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
XX
PT disease.
XX
PS Disclosure; SEQ ID NO 14; 72pp; English.
XX
XX The present invention relates to novel galactosidase (Gal) proteins such
XX
XX as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
XX
XX of the present invention are useful for producing recombinant lysosomal
```

```
enzymes for enzyme replacement therapy for treating human and animal
CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human wild type rGAL-8R protein. This sequence comprises a human rGAL-8
CC protein and a ER retention signal. This sequence is used in the
CC exemplification of the invention.
XX
SQ Sequence 427 AA;
Query Match 100.0%; Score 2314; DB 8; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPGLHGCALALRELALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
Db 1 MOLRNPGLHGCALALRELALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
QY 61 DSCISEKLFMEWMAELMVSEGKADAGEYELICIDDCWMAFQDSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEWMAELMVSEGKADAGEYELICIDDCWMAFQDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLGLGIYADVGNKTCAGPFGSGFYDYDAQTADFAGVDLLKFDGICYDSLENL 180
Db 121 ANYVHSKGLGLGIYADVGNKTCAGPFGSGFYDYDAQTADFAGVDLLKFDGICYDSLENL 180
QY 181 ADGYKHSALNRTGRSIVVSCWPLNMPQKPNYTEIROYCNHWRNPFADIDDSWSKSIK 240
Db 181 ADGYKHSALNRTGRSIVVSCWPLNMPQKPNYTEIROYCNHWRNPFADIDDSWSKSIK 240
QY 241 SILDWTSTFQRIIVDVAGGGWDPMLVIGNFGLSWQVVTQMALWAIMAAPLPMNDL 300
Db 241 SILDWTSTFQRIIVDVAGGGWDPMLVIGNFGLSWQVVTQMALWAIMAAPLPMNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWEPRLSGLAWAVAMINRQIIG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWEPRLSGLAWAVAMINRQIIG 360
QY 361 GPRSYTIAVASIGKGVACNPACFIITQLLPVKRKLGFYEWTSLRSHINPTGTVLQLENT 420
Db 361 GPRSYTIAVASIGKGVACNPACFIITQLLPVKRKLGFYEWTSLRSHINPTGTVLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 12
ADU66918
ID ADU66918 standard; protein; 427 AA.
XX
AC ADU66918;
XX
DT 10-FEB-2005 (first entry)
XX
DE Human alpha-galactosidase protein #6.
XX
KW Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human.
XX
OS Homo sapiens.
XX
PN US2004234516-A1.
XX
PD 25-NOV-2004.
XX
PF 21-MAY-2004; 2004US-00851388.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PR 13-NOV-2001; 2001US-00993059.
XX
PR 20-MAR-2002; 2002US-00103327.
```

PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2004-821274/81.
XX N-PSDB; ADU66917.
XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.
XX
XX Disclosure; SEQ ID NO 14; 88pp; English.
XX
XX The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein.
XX
XX Sequence 427 AA;
Query Match 100.0%; Score 2314; DB 8; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGRLARTPTMGWLHWRFCNLDCCQSEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGRLARTPTMGWLHWRFCNLDCCQSEP 60
Qy 61 DSCISEKLFMEAMELMWSEGWKDAYEYLCIDDCWMAQORSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAMELMWSEGWKDAYEYLCIDDCWMAQORSEGRLOADPQRFPHGIRQL 120
Qy 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYVDIDAQTFADWGVLLKFDGVCYCDLSLENL 180
Db 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYVDIDAQTFADWGVLLKFDGVCYCDLSLENL 180
Qy 181 ADGYKMSLALNRGSRIVSYSCWPLYMWPQKPNYTIROYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKMSLALNRGSRIVSYSCWPLYMWPQKPNYTIROYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSPNQERIYDVAGPGGNDPDMLVGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSPNQERIYDVAGPGGNDPDMLVGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Qy 361 GPRSTYIAVSLGKGVACNPACFITQLLPVKKLGFYEWTSLRSHINPTGTVLLOLENT 420
Db 361 GPRSTYIAVSLGKGVACNPACFITQLLPVKKLGFYEWTSLRSHINPTGTVLLOLENT 420
Qy 421 M 421
Db 421 M 421
RESULT 13
ID AEA27447
XX AEA27447 standard; protein; 427 AA.
XX AC AEA27447;
XX 11-AUG-2005 (first entry)
XX

DE Human alpha-galactosidase protein, rGAL-8R, SEQ ID NO: 14.
XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
KW enzyme.
XX Homo sapiens.
OS
XX US2005125859-A1.
PN
XX 09-JUN-2005.
PD
XX 08-NOV-2004; 2004US-00984389.
PF
XX 26-JUL-2000; 2000US-00626127.
PR
XX 13-NOV-2001; 2001US-00993059.
PR
XX 20-MAR-2002; 2002US-00103327.
XX (LARG-) LARGE SCALE BIOLOGY CORP.
PA Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2005-404004/41.
DR N-PSDB; AEA27446.
DR
XX New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.
XX
XX Disclosure; SEQ ID NO 14; 88pp; English.
FS
XX The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase (gal) enzymes having a post-translational modification
CC provided by the plant expression system. The invention is useful in
CC enzyme replacement therapy for treating lysosomal storage diseases such
CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein.
XX
XX Sequence 427 AA;
Query Match 100.0%; Score 2314; DB 9; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGRLARTPTMGWLHWRFCNLDCCQSEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGRLARTPTMGWLHWRFCNLDCCQSEP 60
Qy 61 DSCISEKLFMEAMELMWSEGWKDAYEYLCIDDCWMAQORSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAMELMWSEGWKDAYEYLCIDDCWMAQORSEGRLOADPQRFPHGIRQL 120
Qy 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYVDIDAQTFADWGVLLKFDGVCYCDLSLENL 180
Db 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYVDIDAQTFADWGVLLKFDGVCYCDLSLENL 180
Qy 181 ADGYKMSLALNRGSRIVSYSCWPLYMWPQKPNYTIROYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKMSLALNRGSRIVSYSCWPLYMWPQKPNYTIROYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSPNQERIYDVAGPGGNDPDMLVGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSPNQERIYDVAGPGGNDPDMLVGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIAINQDPLGKQGYLROQGNFVWERPLSGLAWAVAMINRQSIG 360

Db 301 RHISPOKALLQDKVIAINQDPLCKQGYQLRQGDNFVWFRPLSGLAWAVAMINRQBIG 360
QY 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
Db 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
QY 421 M 421
Db 421 M 421
RESULT 14
ID AAR53765 standard; protein; 429 AA.
XX AAR53765;
XX 25-MAR-2003 (revised)
DT 15-NOV-1994 (first entry)
XX Alpha-galactosidase A.
DE Alpha-galactosidase A.
XX Alpha-galactosidase A; Fabry disease; blood group O; blood group B; CHO;
KW Chinese hamster ovary; enzyme replacement therapy; glycoconjugate;
KW alpha-Gal A.
XX Homo sapiens.
OS Homo sapiens.
XX Key Location/Qualifiers
FH Modified-site 236
FT /note= "putative N-glycosylation site"
FT Modified-site 305
FT /note= "putative N-glycosylation site"
FT Modified-site 312
FT /note= "putative N-glycosylation site"
FT Modified-site 345
FT /note= "putative N-glycosylation site"
XX WO9412628-A1.
FN 09-JUN-1994.
XX 30-NOV-1993; 93WO-US011539.
XX 30-NOV-1992; 92US-00983451.
XX (MOUN) MOUNT SINAI SCHOOL MEDICINE.
XX Desnick RJ, Bishop DF, Ioannou VA;
XX WPI; 1994-200257/24.
DR N-PSDB; AAQ66241.
XX Prodn of human alpha-galactosidase A - by culturing cells contg the
PT coding sequence and the beta-galactosidase alpha-2,6-sialyl transferase
PT gene and regulation sequences.
XX Disclosure; Page 102-103; 156pp; English.
XX The cDNA sequence (AAQ66241) of human alpha-galactosidase A (alpha-Gal A)
CC was determined. The predicted amino acid sequence (AAR53765) showed 50%
CC homology with human alpha-N-acetylgalactosaminidase (AAR53766) (alpha-
CC galactosidase B) and showed short regions of homology with yeast Mel 1
CC (AAR53767). Recombinant alpha-Gal A was produced in CHO cells. The enzyme
CC can be used for Fabry disease enzyme replacement therapy, to convert
CC blood group B to O, or to hydrolyze alpha-D-galactosyl residues from
CC glycoconjugates. (Updated on 25-MAR-2003 to correct PN field.)
XX Sequence 429 AA;
SQ
Query Match 100.0%; Score 2314; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;

Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPEHLGCALALREFALALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCQEP 60
Db 1 MOLRNPEHLGCALALREFALALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCQEP 60
QY 61 DSCISEKLFMEWAEIWMVSEGKWDAGYEYLCIDDCWMAFQDSEGRSQADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEWAEIWMVSEGKWDAGYEYLCIDDCWMAFQDSEGRSQADPQRFPHGIRQL 120
QY 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYDYDAQTADFAGWDVLLKFDGCYCDLSLENL 180
Db 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYDYDAQTADFAGWDVLLKFDGCYCDLSLENL 180
QY 181 ADGKHSALNLRGTSIVYSCWPLYMWPQKPNYTEIROVCHNRNPFADIDDSWKSIIK 240
Db 181 ADGKHSALNLRGTSIVYSCWPLYMWPQKPNYTEIROVCHNRNPFADIDDSWKSIIK 240
QY 241 SILDWTSEFQRIVDVAGPGWMDPDLVIGNFGLSMNQVTTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSEFQRIVDVAGPGWMDPDLVIGNFGLSMNQVTTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKVIAINQDPLCKQGYQLRQGDNFVWFRPLSGLAWAVAMINRQBIG 360
Db 301 RHISPOKALLQDKVIAINQDPLCKQGYQLRQGDNFVWFRPLSGLAWAVAMINRQBIG 360
QY 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
Db 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
QY 421 M 421
Db 421 M 421
RESULT 15
ID AAR63234 standard; protein; 429 AA.
XX AAR63234;
XX 29-JUN-1995 (first entry)
XX Human alpha-galactosidase A protein.
DE Human; alpha-galactosidase A; expression; mammalian; expression plasmid;
KW chromatography; fusion protein; Staphylococcus; protein A; cleavage site;
KW housekeeping gene; collagenase; affinity chromatography; galactosidase;
KW enzyme replacement therapy; gene therapy; lysosomal storage disorder;
KW Fabry disease; alpha-D-galacto-glycoconjugate; blood group.
OS Homo sapiens.
XX Key Location/Qualifiers
FH Modified-site 139..141
FT /note= "N-linked glycosylation"
FT Modified-site 192..194
FT /note= "N-linked glycosylation"
FT Modified-site 215..217
FT /note= "N-linked glycosylation"
FT Modified-site 408..410
FT /note= "N-linked glycosylation"
XX US5356804-A.
PN 18-OCT-1994.
PD 24-OCT-1990; 90US-00602824.
PF 24-OCT-1990; 90US-00602824.
XX 24-OCT-1990; 90US-00602824.
XX (MOUN) MOUNT SINAI SCHOOL MEDICINE.
PA
XX

PI Ioannou YA, Desnick RJ, Bishop DF;

XX WPI; 1994-340330/42.
DR N-PSDB; AAQ77861.

XX Recombinant human alpha-galactosidase A prodn. - using a mammalian host
PT cell expression system to obtain high yields of enzymatically active
PT enzyme.

XX Claim 3; Col 43-46; 60pp; English.

XX The amino acid sequence of the human alpha-galactosidase A (hAGA). The
CC gene encodes a protein of 429 amino acids. Methods to overexpress and
CC purify the protein include: (i) placing the 1.45 Kb human hAGA gene
CC fragment in a mammalian expression plasmid e.g. AAP91023 and purifying
CC total cell protein by chromatographic methods including affinity
CC chromatography or (ii) by producing a fusion protein comprising hAGA
CC fused to easily purified protein sequences e.g. the Staphylococcus protein
CC A, with cleavage site for a "housekeeping" gene e.g. collagenase cleavage
CC site, inserted between the two sequences in a mammalian expression
CC plasmid. The fusion protein can be easily purified by affinity
CC chromatography and the galactosidase protein released from the fusion
CC protein whilst still bound to the column by cleavage with collagenase.
CC The methods result in active alpha-galactosidase A protein. The hAGA can
CC be used for enzyme replacement therapy in patients with the lysosomal
CC storage disorder, Fabry disease. The protein can also be used in vitro to
CC modify alpha-D-galacto-glycoconjugates in a variety of processes e.g.
CC converting blood group B erythrocytes to group O

XX Sequence 429 AA;

Query Match 100.0%; Score 2314; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQLRNPGLHGCALALRPLALVSDIPGARALDNGLARTPTMGWLHWRFNCLDCQEEP 60
Db 1 MQLRNPGLHGCALALRPLALVSDIPGARALDNGLARTPTMGWLHWRFNCLDCQEEP 60
Qy 61 DSCISEKLFMEAEIWMVSEGWKADGYEYLCIDDCWAPQDSEGRLOADPQPFPHGIROL 120
Db 61 DSCISEKLFMEAEIWMVSEGWKADGYEYLCIDDCWAPQDSEGRLOADPQPFPHGIROL 120
Qy 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYDYIDAQTFADWGVLLKFDGICYCDLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYDYIDAQTFADWGVLLKFDGICYCDLENL 180
Qy 181 ADGYKMSIALNRTGRSIVYSCWPLYNWPFQKPNYTEIRQYCNHWRNFADIDDSWKSIX 240
Db 181 ADGYKMSIALNRTGRSIVYSCWPLYNWPFQKPNYTEIRQYCNHWRNFADIDDSWKSIX 240
Qy 241 SILDWTSPNQBRIVDVAGPGWNDPDLVIGNFLSWNQVQTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSPNQBRIVDVAGPGWNDPDLVIGNFLSWNQVQTQMALWAIMAAPLFMSNDL 300
Qy 301 RHISFQAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVEWRPLSGLAWAVAMINRQBIG 360
Db 301 RHISFQAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVEWRPLSGLAWAVAMINRQBIG 360
Qy 361 GPRSYTIAVSLGKGVACNPACFTQLLPVKRKLGFYEWTSRLSRSHINPTGTVLQLENT 420
Db 361 GPRSYTIAVSLGKGVACNPACFTQLLPVKRKLGFYEWTSRLSRSHINPTGTVLQLENT 420
Qy 421 M 421
Db 421 M 421

Search completed: December 24, 2005, 08:54:18
Job time : 189 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 08:54:29 ; Search time 166 Seconds
(without alignments)
1059.676 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MOLRNPGLGALALRFLA.....RLRSHINPTGTVLLQLENTM 421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2314	100.0	421	US-09-993-059-12 Sequence 12, Appl
2	2314	100.0	421	US-10-103-327-12 Sequence 12, Appl
3	2314	100.0	421	US-10-602-219-12 Sequence 12, Appl
4	2314	100.0	421	US-10-602-220-12 Sequence 12, Appl
5	2314	100.0	421	US-10-851-388-12 Sequence 12, Appl
6	2314	100.0	421	US-10-984-389-12 Sequence 12, Appl
7	2314	100.0	427	US-09-993-059-14 Sequence 14, Appl
8	2314	100.0	427	US-10-103-327-14 Sequence 14, Appl
9	2314	100.0	427	US-10-602-219-14 Sequence 14, Appl
10	2314	100.0	427	US-10-602-220-14 Sequence 14, Appl
11	2314	100.0	427	US-10-851-388-14 Sequence 14, Appl
12	2314	100.0	427	US-10-984-389-14 Sequence 14, Appl
13	2314	100.0	429	US-10-411-037-68 Sequence 68, Appl
14	2314	100.0	429	US-10-411-026-68 Sequence 68, Appl
15	2314	100.0	429	US-10-410-962-68 Sequence 68, Appl
16	2314	100.0	429	US-10-410-049-68 Sequence 68, Appl
17	2314	100.0	429	US-10-410-930-68 Sequence 68, Appl
18	2314	100.0	429	US-10-410-997-68 Sequence 68, Appl
19	2314	100.0	429	US-10-411-012-68 Sequence 68, Appl
20	2314	100.0	429	US-10-410-913-68 Sequence 68, Appl
21	2314	100.0	429	US-10-370-715B-278 Sequence 278, App
22	2314	100.0	429	US-10-410-980-68 Sequence 68, Appl
23	2314	100.0	429	US-10-410-897-68 Sequence 68, Appl
24	2314	100.0	431	US-09-993-059-10 Sequence 10, Appl
25	2314	100.0	431	US-10-103-327-10 Sequence 10, Appl
26	2314	100.0	431	US-10-602-219-10 Sequence 10, Appl
27	2314	100.0	431	US-10-602-220-10 Sequence 10, Appl

28	2314	100.0	431	US-10-851-388-10	Sequence 10, Appl
29	2314	100.0	431	US-10-984-389-10	Sequence 10, Appl
30	2314	100.0	435	US-09-993-059-6	Sequence 6, Appl
31	2314	100.0	435	US-10-103-327-6	Sequence 6, Appl
32	2314	100.0	435	US-10-602-219-6	Sequence 6, Appl
33	2314	100.0	435	US-10-602-220-6	Sequence 6, Appl
34	2314	100.0	435	US-10-851-388-6	Sequence 6, Appl
35	2314	100.0	435	US-10-984-389-6	Sequence 6, Appl
36	2314	100.0	478	US-10-131-410-106	Sequence 106, App
37	2309	99.8	424	US-09-993-059-8	Sequence 8, Appl
38	2309	99.8	424	US-10-103-327-8	Sequence 8, Appl
39	2309	99.8	424	US-10-602-219-8	Sequence 8, Appl
40	2309	99.8	424	US-10-602-220-8	Sequence 8, Appl
41	2309	99.8	424	US-10-851-388-8	Sequence 8, Appl
42	2309	99.8	424	US-10-984-389-8	Sequence 8, Appl
43	2309	99.8	428	US-09-993-059-4	Sequence 4, Appl
44	2309	99.8	428	US-10-103-327-4	Sequence 4, Appl
45	2309	99.8	428	US-10-602-219-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-12
; Sequence 12, Application US/09993059
; Publication No. US2002008024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-12

Query Match	100.0%	Score 2314;	DB 3;	Length 421;
Best Local Similarity	100.0%	Pred. No. 1.7e-221;		
Matches 421;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MOLRNPGLGALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCMNCDCQBP	60	
Db	1	MOLRNPGLGALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCMNCDCQBP	60	
Qy	61	DSCISEKLFMEAEIWMYSEGWKDAYEYLCIDDCWAPQDSRSEGLQADPQFPFGIRQL	120	
Db	61	DSCISEKLFMEAEIWMYSEGWKDAYEYLCIDDCWAPQDSRSEGLQADPQFPFGIRQL	120	
Qy	121	ANYVHSGKLGLGIYADVGNKTCAGFPGSFGYYDIDAQTFADGWVLLKFDGCGYCSLENL	180	
Db	121	ANYVHSGKLGLGIYADVGNKTCAGFPGSFGYYDIDAQTFADGWVLLKFDGCGYCSLENL	180	
Qy	181	ADGYKMSLALNRTGRSIVYCEWPLYNWPQKNYTEIRQYCNHWRNFADIDDSWKSITK	240	
Db	181	ADGYKMSLALNRTGRSIVYCEWPLYNWPQKNYTEIRQYCNHWRNFADIDDSWKSITK	240	
Qy	241	SILDWTSPNQRIVDVAGPGGNDPDLVIGNGLSNQOVTOMALWAIMAPLFWMSNDL	300	
Db	241	SILDWTSPNQRIVDVAGPGGNDPDLVIGNGLSNQOVTOMALWAIMAPLFWMSNDL	300	
Qy	301	RHISFOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWFRPLSGLAWAMINRQIG	360	
Db	301	RHISFOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWFRPLSGLAWAMINRQIG	360	
Qy	361	GPRSYTTIIVASLKGKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT	420	

Db 361 GPRSYTIAVSLGKGVACNPAFCFITQLLPVKKLGFEYWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 2

US-10-103-327-12
; Sequence 12, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSGENETIC EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-12

Query Match 100.0%; Score 2314; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221; Indels 0; Gaps 0;
Matches 421; Conservative 0; Mismatches 0;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPODSEGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPODSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGYGYDDAQTTFADWGVLLKFDGVCDSLENTL 180
Db 121 ANYVHSKGLGIYADVGNKTCAGPFGSGYGYDDAQTTFADWGVLLKFDGVCDSLENTL 180
QY 181 ADGKXMSLALNRTGSRIVYSCWPLYMWPFOKPNYTEIROQCNHWRNPADIDDSWKSITK 240
Db 181 ADGKXMSLALNRTGSRIVYSCWPLYMWPFOKPNYTEIROQCNHWRNPADIDDSWKSITK 240
QY 241 SILDWTSTFNOERIVDVAGPGGNDPDMLVIGNFGLSWNQVVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFNOERIVDVAGPGGNDPDMLVIGNFGLSWNQVVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVSLGKGVACNPAFCFITQLLPVKKLGFEYWTSLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVSLGKGVACNPAFCFITQLLPVKKLGFEYWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 3

US-10-602-219-12
; Sequence 12, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:

; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIGENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-219-12

Query Match 100.0%; Score 2314; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221; Indels 0; Gaps 0;
Matches 421; Conservative 0; Mismatches 0;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPODSEGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPODSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGYGYDDAQTTFADWGVLLKFDGVCDSLENTL 180
Db 121 ANYVHSKGLGIYADVGNKTCAGPFGSGYGYDDAQTTFADWGVLLKFDGVCDSLENTL 180
QY 181 ADGKXMSLALNRTGSRIVYSCWPLYMWPFOKPNYTEIROQCNHWRNPADIDDSWKSITK 240
Db 181 ADGKXMSLALNRTGSRIVYSCWPLYMWPFOKPNYTEIROQCNHWRNPADIDDSWKSITK 240
QY 241 SILDWTSTFNOERIVDVAGPGGNDPDMLVIGNFGLSWNQVVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFNOERIVDVAGPGGNDPDMLVIGNFGLSWNQVVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVSLGKGVACNPAFCFITQLLPVKKLGFEYWTSLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVSLGKGVACNPAFCFITQLLPVKKLGFEYWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 4
US-10-602-220-12
; Sequence 12, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-220-12

Query Match 100.0%; Score 2314; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221; Mismatches 0; Indels 0; Gaps 0;
Matches 421; Conservative 0;
QY 1 MOLRNPHELGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPHELGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYDDAQTADGWVLLKFDGCGYCDLSLENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYDDAQTADGWVLLKFDGCGYCDLSLENL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKSJK 240
DB 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKSJK 240
QY 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQIEG 360
DB 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQIEG 360
QY 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420

QY 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421
RESULT 5
US-10-851-388-12
; Sequence 12, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: TURPEN, Stephen A.
; APPLICANT: GARGER, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-851-388-12

Query Match 100.0%; Score 2314; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPHELGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYDDAQTADGWVLLKFDGCGYCDLSLENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYDDAQTADGWVLLKFDGCGYCDLSLENL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKSJK 240
DB 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKSJK 240
QY 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQIEG 360
DB 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQIEG 360
QY 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 6
US-10-984-389-12
; Sequence 12, Application US/10984389

```
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-389-12

Query Match      100.0%; Score 2314; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
DB      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60

QY      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
DB      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120

QY      121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
DB      121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180

QY      181 ADGYKHSALNRTGRSIVYCEWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWSKSIK 240
DB      181 ADGYKHSALNRTGRSIVYCEWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWSKSIK 240

QY      241 SILDWTSPNQERIYDVAGPGWNPDMVLVIGNFGLSWNQVVTOMALWAIMAAPLFMSNDL 300
DB      241 SILDWTSPNQERIYDVAGPGWNPDMVLVIGNFGLSWNQVVTOMALWAIMAAPLFMSNDL 300

QY      301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQETG 360
DB      301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQETG 360

QY      361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
DB      361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420

QY      421 M 421
DB      421 M 421

RESULT 7
US-09-993-059-14
; Sequence 14, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14

Query Match      100.0%; Score 2314; DB 4; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
DB      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60

QY      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
DB      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120

QY      121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
DB      121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180

QY      181 ADGYKHSALNRTGRSIVYCEWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWSKSIK 240
DB      181 ADGYKHSALNRTGRSIVYCEWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWSKSIK 240

QY      241 SILDWTSPNQERIYDVAGPGWNPDMVLVIGNFGLSWNQVVTOMALWAIMAAPLFMSNDL 300
DB      241 SILDWTSPNQERIYDVAGPGWNPDMVLVIGNFGLSWNQVVTOMALWAIMAAPLFMSNDL 300

QY      301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQETG 360
DB      301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQETG 360

QY      361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
DB      361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420

QY      421 M 421
DB      421 M 421

RESULT 8
US-10-103-327-14
; Sequence 14, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-14

Query Match      100.0%; Score 2314; DB 4; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
DB      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60

QY      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
DB      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120

QY      121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
DB      121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180

QY      181 ADGYKHSALNRTGRSIVYCEWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWSKSIK 240
DB      181 ADGYKHSALNRTGRSIVYCEWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWSKSIK 240

QY      241 SILDWTSPNQERIYDVAGPGWNPDMVLVIGNFGLSWNQVVTOMALWAIMAAPLFMSNDL 300
DB      241 SILDWTSPNQERIYDVAGPGWNPDMVLVIGNFGLSWNQVVTOMALWAIMAAPLFMSNDL 300

QY      301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQETG 360
DB      301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQETG 360

QY      361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
DB      361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420

QY      421 M 421
DB      421 M 421
```

Qy 61 DSCISEKLFMEAMBLMVSEGWDAGYELVCIDDCWMAPOQSEGRLOADPQRPFGHIGIROL 120
Db 61 DSCISEKLFMEAMBLMVSEGWDAGYELVCIDDCWMAPOQSEGRLOADPQRPFGHIGIROL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
Qy 181 ADGKXMSLALNRTGRSIVYSCWPLVMFPQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLVMFPQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSENOBRIYDVAGPGGNDPDMVLVGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSENOBRIYDVAGPGGNDPDMVLVGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKDVIAINODPLGKQGYLROQGNFEVWERPLSGLAWAVAMINRQIG 360
Db 301 RHISPOKALLQDKDVIAINODPLGKQGYLROQGNFEVWERPLSGLAWAVAMINRQIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
Qy 421 M 421
Db 421 M 421

RESULT 9

US-10-602-219-14
; Sequence 14, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-219-14

Query Match 100.0%; Score 2314; DB 4; Length 427;

Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLNPELHGLCALALRFLALVSWDIFGARALDNLGLARTPTMGWLHWRFCNLDQCEP 60
Db 1 MOLNPELHGLCALALRFLALVSWDIFGARALDNLGLARTPTMGWLHWRFCNLDQCEP 60
Qy 61 DSCISEKLFMEAMBLMVSEGWDAGYELVCIDDCWMAPOQSEGRLOADPQRPFGHIGIROL 120
Db 61 DSCISEKLFMEAMBLMVSEGWDAGYELVCIDDCWMAPOQSEGRLOADPQRPFGHIGIROL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
Qy 181 ADGKXMSLALNRTGRSIVYSCWPLVMFPQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLVMFPQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSENOBRIYDVAGPGGNDPDMVLVGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSENOBRIYDVAGPGGNDPDMVLVGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKDVIAINODPLGKQGYLROQGNFEVWERPLSGLAWAVAMINRQIG 360
Db 301 RHISPOKALLQDKDVIAINODPLGKQGYLROQGNFEVWERPLSGLAWAVAMINRQIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
Qy 421 M 421
Db 421 M 421

RESULT 10

US-10-602-220-14
; Sequence 14, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-220-14

Query Match      100.0%; Score 2314; DB 4; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQBP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQBP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIQRL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIQRL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADMGVDLLKFDGCGYCDLSN 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADMGVDLLKFDGCGYCDLSN 180
QY 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240
QY 241 SILDWTSTFNOERIYDVAGPGGWNDDMLVIGNFGLSNQOQVTQMALWAIMAAPLFWMSNDL 300
Db 241 SILDWTSTFNOERIYDVAGPGGWNDDMLVIGNFGLSNQOQVTQMALWAIMAAPLFWMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQ 360
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQ 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKKLGFYEWTSRLRSHINPTGTVLLOLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKKLGFYEWTSRLRSHINPTGTVLLOLENT 420
QY 421 M 421
Db 421 M 421

RESULT 12
US-10-984-389-14
; Sequence 14, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-389-14

Query Match      100.0%; Score 2314; DB 5; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQBP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQBP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIQRL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIQRL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADMGVDLLKFDGCGYCDLSN 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADMGVDLLKFDGCGYCDLSN 180
QY 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240

Query Match      100.0%; Score 2314; DB 4; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQBP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQBP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIQRL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIQRL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADMGVDLLKFDGCGYCDLSN 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADMGVDLLKFDGCGYCDLSN 180
QY 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240
QY 241 SILDWTSTFNOERIYDVAGPGGWNDDMLVIGNFGLSNQOQVTQMALWAIMAAPLFWMSNDL 300
Db 241 SILDWTSTFNOERIYDVAGPGGWNDDMLVIGNFGLSNQOQVTQMALWAIMAAPLFWMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQ 360
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQ 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKKLGFYEWTSRLRSHINPTGTVLLOLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKKLGFYEWTSRLRSHINPTGTVLLOLENT 420
QY 421 M 421
Db 421 M 421

RESULT 11
US-10-851-388-14
; Sequence 14, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-851-388-14

Query Match      100.0%; Score 2314; DB 5; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQBP 60
```


QY 241 SILDWTSTFNOERIIVDVAGPGGNDPDMVLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSTFNOERIIVDVAGPGGNDPDMVLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
QY 361 GPRSYTTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 13
US-10-411-037-68
; Sequence 68, Application US/10411037
; Publication No. US20040043446A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Defrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Bove, Caryn
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA
; FILE REFERENCE: 040853-01-5082
; CURRENT APPLICATION NUMBER: US/10/411,037
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-037-68

Query Match 100.0%; Score 2314; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e-221; Indels 0; Gaps 0;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPELHGLCALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCNLDCCQEP 60
DB 1 MOLRNPELHGLCALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCNLDCCQEP 60
QY 61 DSCISEKLFMEAEIWMVSEGWDAGYEYLCIDDCWMAFQDSEGRQADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEIWMVSEGWDAGYEYLCIDDCWMAFQDSEGRQADPQRFPHGIRQL 120
QY 121 ANYVHSKGLKGIYADVGNKTCAGPPSGFYGYDDAQTADFADGWVLLKFDGCYCDLSLENL 180
DB 121 ANYVHSKGLKGIYADVGNKTCAGPPSGFYGYDDAQTADFADGWVLLKFDGCYCDLSLENL 180

QY 181 ADGKXKMSLALNRTGRSIVYSCFPLWMPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKXKMSLALNRTGRSIVYSCFPLWMPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSTFNOERIIVDVAGPGGNDPDMVLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSTFNOERIIVDVAGPGGNDPDMVLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
QY 361 GPRSYTTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 14
US-10-411-026-68
; Sequence 68, Application US/10411026
; Publication No. US20040063911A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Defrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; TITLE OF INVENTION: PROTEIN REMODELING METHODS AND PROTEINS/PEPTIDES PRODUCED BY THE
; FILE REFERENCE: 040853-01-5053
; CURRENT APPLICATION NUMBER: US/10/411,026
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-026-68

Query Match 100.0%; Score 2314; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e-221; Indels 0; Gaps 0;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPELHGLCALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCNLDCCQEP 60
DB 1 MOLRNPELHGLCALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCNLDCCQEP 60
QY 61 DSCISEKLFMEAEIWMVSEGWDAGYEYLCIDDCWMAFQDSEGRQADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEIWMVSEGWDAGYEYLCIDDCWMAFQDSEGRQADPQRFPHGIRQL 120
QY 121 ANYVHSKGLKGIYADVGNKTCAGPPSGFYGYDDAQTADFADGWVLLKFDGCYCDLSLENL 180

Db 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVLLKFDGVCDSLENL 180
Qy 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFNOERIIVDVAGPGGWNDDPMLVIGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIIVDVAGPGGWNDDPMLVIGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQEIIG 360
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQEIIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
Qy 421 M 421
Db 421 M 421

RESULT 15
US-10-410-962-68
; Sequence 68, Application US/10410962
; Publication No. US20040077836A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Defrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Bowe, Caryn
; TITLE OF INVENTION: GLYCOCONJUGATION OF G-CSF
; FILE REFERENCE: 040853-01-5054
; CURRENT APPLICATION NUMBER: US/10/410,962
; PRIOR FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-410-962-68

Query Match 100.0%; Score 2314; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLRNPDLHGICALALRFLALYSWDIPGARALDNLGARTPTMGWLHWRFCMNLDCQEEP 60
Db 1 MOLRNPDLHGICALALRFLALYSWDIPGARALDNLGARTPTMGWLHWRFCMNLDCQEEP 60
Qy 61 DSCISEKLFMEAEMLMVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADPQRPFGIROL 120
Db 61 DSCISEKLFMEAEMLMVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADPQRPFGIROL 120

Qy 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVLLKFDGVCDSLENL 180
Db 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVLLKFDGVCDSLENL 180
Qy 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFNOERIIVDVAGPGGWNDDPMLVIGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIIVDVAGPGGWNDDPMLVIGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQEIIG 360
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQEIIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
Qy 421 M 421
Db 421 M 421

Search completed: December 24, 2005, 09:11:10
Job time : 168 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 08:50:53 ; Search time 13 Seconds
(without alignments)
230.993 Million cell updates/sec

Title: US-10-602-219-12

Perfect score: 2314

Sequence: 1 MQLRNPBLHGLCALALRFLA.....RLRSHINPTGVLLQLENTM 421

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 54001 seqs, 7132810 residues

Total number of hits satisfying chosen parameters: 54001

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA New:

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	3.7	437	7	US-11-073-626-3
2	82	3.5	437	7	US-11-073-626-1
3	81	3.5	310	7	US-11-074-176-108
4	81	3.5	944	7	US-11-057-058-68
5	80.5	3.5	414	6	US-10-467-657-32
6	80.5	3.5	414	6	US-10-467-657-8372
7	80	3.5	861	7	US-11-038-284-36
8	79.5	3.4	1827	7	US-11-057-058-62
9	79	3.4	322	6	US-10-467-657-8554
10	79	3.4	368	7	US-11-129-143-62
11	78.5	3.4	622	6	US-10-510-386-236
12	78	3.4	760	6	US-10-821-234-1141
13	78	3.4	786	6	US-10-467-962B-103
14	77.5	3.3	522	6	US-10-131-826A-450
15	77.5	3.3	903	7	US-11-057-058-65
16	77	3.3	734	6	US-10-652-893-2
17	76.5	3.3	320	6	US-10-467-657-3254
18	76.5	3.3	717	6	US-10-793-626-3022
19	76	3.3	2769	7	US-11-113-424-14
20	75.5	3.3	275	7	US-11-055-822-2
21	75.5	3.3	843	6	US-10-645-441-8
22	75	3.2	402	6	US-10-650-326B-21
23	75	3.2	402	6	US-11-051-568-29
24	75	3.2	627	6	US-10-873-528-191
25	74.5	3.2	482	6	US-10-793-626-1314

26	74.5	3.2	573	6	US-10-467-657-684	Sequence 684, App
27	74.5	3.2	966	7	US-11-057-058-67	Sequence 67, Appl
28	74	3.2	417	6	US-10-467-657-6612	Sequence 6612, Ap
29	73.5	3.2	932	7	US-11-057-058-59	Sequence 59, Appl
30	73	3.2	514	7	US-11-103-037-3	Sequence 3, Appl
31	73	3.2	734	7	US-11-137-465-65	Sequence 65, Appl
32	72	3.1	515	6	US-10-630-203-6	Sequence 6, Appl
33	72	3.1	842	7	US-11-038-284-38	Sequence 38, Appl
34	72	3.1	847	7	US-11-038-284-42	Sequence 42, Appl
35	70.5	3.0	316	6	US-10-821-234-1526	Sequence 1526, Ap
36	70.5	3.0	834	6	US-10-928-446A-188	Sequence 188, App
37	70.5	3.0	834	6	US-10-928-446A-192	Sequence 192, App
38	70.5	3.0	834	6	US-10-928-446A-194	Sequence 194, App
39	70.5	3.0	834	6	US-10-928-446A-196	Sequence 196, App
40	70.5	3.0	834	6	US-10-928-446A-198	Sequence 198, App
41	70.5	3.0	834	6	US-10-928-446A-200	Sequence 200, App
42	70.5	3.0	834	6	US-10-928-446A-202	Sequence 202, App
43	70.5	3.0	942	6	US-10-928-446A-190	Sequence 190, App
44	70.5	3.0	947	6	US-10-928-446A-182	Sequence 182, App
45	70.5	3.0	950	6	US-10-928-446A-184	Sequence 184, App

ALIGNMENTS

RESULT 1

US-11-073-626-3
; Sequence 3, Application US/11073626
; Publication No. US20050244926A1
; GENERAL INFORMATION:

; APPLICANT: KUROSAWA, KEIKO
; APPLICANT: HIROKAWA, KOZO
; APPLICANT: KAJIYAMA, NAOKI
; TITLE OF INVENTION: Novel fructosyl peptide oxidase
; FILE REFERENCE: 227590US0
; CURRENT APPLICATION NUMBER: US/11/073,626
; PRIOR FILING DATE: 2005-03-08
; PRIOR APPLICATION NUMBER: US/10/232,655
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Eupenicillium terrenum
US-11-073-626-3

Query Match 3.7%; Score 85; DB 7; Length 437;

Best Local Similarity 20.4%; Pred. No. 0.62;
Matches 46; Conservative 29; Mismatches 71; Indels 80; Gaps 12;

Qy	32	LDNGLRPTMGWLHWRFCNLDQCPEPDCISEKLFMEAEALMVSEGWKDAGYEYLCI	91
Db	115	LDAGIGLEKTNVWL-----ESEDILAKAPNFTREQV---KGNKG-----LFCT	155
Qy	92	DDCNMAPORDEGRLOADPQRFPHGIROLANYVHSGKGLKG-----IYADVGNKT	141
Db	156	DGGWLAARK-----AINAIGIFLDQKGVKFGGAGTQQPLFAADG-KT	199
Qy	142	CAGFGSGGYVDIDAQTAD-----WGVDLKF-DGC-----YCDSENLADGY	184
Db	200	CIGLETTDG-----TKYFADKVLAAAGAWSPTLDLEDQCYSKAWVFAHIQUTPK	254
Qy	185	KHMSIALNRTGRSIVYSCENPLYMFPKPNVTEIROYCNHWRNFA	230
Db	255	KNV-----PVYDGEYGF-----PFEPNEYGVIKVCDEFPFGFS	287

RESULT 2

US-11-073-626-1
; Sequence 1, Application US/11073626
; Publication No. US20050244926A1
; GENERAL INFORMATION:
; APPLICANT: KUROKAWA, KEIKO
; APPLICANT: HIROKAWA, KOZO
; APPLICANT: KAJIYAMA, NAOKI
; TITLE OF INVENTION: Novel fructosyl peptide oxidase
; FILE REFERENCE: 227590050
; CURRENT APPLICATION NUMBER: US/11/073,626
; CURRENT FILING DATE: 2005-03-08
; PRIOR APPLICATION NUMBER: US/10/232,655
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Coniochaeta sp.
US-11-073-626-1

Query Match 3.5%; Score 82; DB 7; Length 437;

Best Local Similarity 19.5%; Pred. No. 1.2;
Matches 42; Conservative 31; Mismatches 82; Indels 60; Gaps 10;
QY 33 DNLGARTPMGLHWRFCNLDCEBPDSCISEKLFMEAEMLVSEGWKADGYEYL CID 92
DB 116 DAGAGLEKTHAWL-----DNEDEILSKMPLQLQDQI---OQWKAIWSQ---D 156
QY 93 DCWMAPOQDSEGRQLADPQRPFGHRLQRLAVVHSGKLGKLG-----IYADVGNKTC 142
DB 157 GGLWLAAK-----ALNAGQFLKRGVGFGGAGSFQKPLFDDGEGT-TC 200
QY 143 AGFPGSFG---YYIDAQTFADWGVLLKFDGCYCDLSLENLADGYKHMSL-----ALNRTG 195
DB 201 IGVETADGPKYADKVLVAAGAWSPTLVDLQDCCSK---AWVYAHQLTPEEAAYKG 256
QY 196 RSIVYSCWPLVWPFQKPNYTEIRQYCNHWRNFA 230
DB 257 VPVVYNGEFGF----FFEFNEFGVIKVCDEFPFGFS 287

RESULT 3

US-11-074-176-108
; Sequence 108, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klienhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 108
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus

US-11-074-176-108

Query Match 3.5%; Score 81; DB 7; Length 310;
Best Local Similarity 20.9%; Pred. No. 0.98;
Matches 54; Conservative 32; Mismatches 114; Indels 58; Gaps 10;
QY 56 QCEPDPDSCISEKLFMEAEMLVSEGWKADGYEYL CIDDCWMAPOQDSEGRQLADPQRPFGH 115
DB 10 CYNEEESI---PLFYNAVE-KVMQTILDLDEYFWFVNDGSSDKSLEENRQLQKNDPEHVH 65
QY 116 GRLQRLAVVHSGKLGKLIYADVGNKTCAGFPFGSGFYDYIDAQTFADWGVLLKFDGCYCD 175
DB 66 YVFSRNFGEAAALYAGLQASIGDYIVV-----MDVDLQ-----D 100
QY 176 SLENLADGYKHMSLA-----LNRGTSIVVSCWPLVWPFQKPNYTEIRQYCNHWRNFA 222
DB 101 PPEFLFEMYNITKGYDCIGTRRVDRKGEAKFKSFLSDMFYKLVNKLISNTEIVSGARDY 160
QY 223 CNHWRNFA-----IDDSWKSISILDWTSF-----NOERIVDVAGFGGWNDPMDLVI 270
DB 161 RMTQMVDAVLSMPEYRNFSGKIFSWGVGFKTKYLDYHNVHVER---VAGESDNTWKLKFKY 217
QY 271 GNFGLS-WNQVQTQMALW 287
DB 218 AMDGIADPSQAPLNIAVW 235

RESULT 4

US-11-057-058-68
; Sequence 68, Application US/11057058
; Publication No. US2005024400A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan
; APPLICANT: Maga, John
; TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
; FILE REFERENCE: SYN-011
; CURRENT APPLICATION NUMBER: US/11/057,058
; CURRENT FILING DATE: 2005-02-10
; PRIOR APPLICATION NUMBER: US 60/543,812
; PRIOR FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-057-058-68

Query Match 3.5%; Score 81; DB 7; Length 944;

Best Local Similarity 26.5%; Pred. No. 4.3;
Matches 36; Conservative 13; Mismatches 49; Indels 38; Gaps 7;
QY 75 LMVSEGWKADGYEYL CIDDCWMAPOQDSEGRQLADPQRPFGH-----IRQLAN----- 122
DB 410 LEVDQGPDD---HNLPCDVIWLDIEHADGKRYFTWDPSPFPQPTMLERLASKRRKLVAI 466
QY 123 -----YVHSGKLGKLIYADVGNKTCAG-----PFGSGFYDYIDAQTFADWGV 164
DB 467 VDPHIKVDGSGYRVHVELRNGLYV-----KTRDGSDEYGCWPGSAGYPDFTNPTWAWWA 522
QY 165 DLLKFDGCYCDLSLENL 180
DB 523 NMFSYDN-YEGSAPNL 537

RESULT 5

US-10-467-657-32
; Sequence 32, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia

```

; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 32
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-32

Query Match 3.5%; Score 80.5; DB 6; Length 414;
Best Local Similarity 19.4%; Pred. No. 1.6;
Matches 37; Conservative 30; Mismatches 79; Indels 45; Gaps 7;

Qy 17 RFLALVWDIPGARALDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELM 76
Db 264 RFREIIGNRLAELNRIGNAFEDTPI-----EVLILLVIDKDKTDGGCIRELYDCKTDTLL 317
Qy 77 VSEGWKAGYEYLCIDDCWAPORSEGRLOADPQRFPHGIROLANYVHSGKLGIYAD 136
Db 318 AADTWQ-----IBPDLWQTVEPAPPKKEDPVILLEHRCRDAAAKRIARELR---FSK 367
Qy 137 VGNKTCAGFGSGYDIDAQTFADMGVDLLKFKGCGYCDLSLENLADGKYKMSLALNR 196
Db 368 MVNE-IEGWPHA-----EFDG-FCDLNCLTQAEKY-----GK 398
Qy 197 SIVVSCSEMPLY 207
Db 399 KHYFPCSLPLF 409

RESULT 6
US-10-467-657-8372
; Sequence 8372, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 8372
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8372

Query Match 3.5%; Score 80.5; DB 6; Length 414;
Best Local Similarity 19.4%; Pred. No. 1.6;
Matches 37; Conservative 30; Mismatches 79; Indels 45; Gaps 7;

Qy 17 RFLALVWDIPGARALDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELM 76
Db 264 RFREIIGNRLAELNRIGNAFEDTPI-----EVLILLVIDKDKTDGGCIRELYDCKTDTLL 317
Qy 77 VSEGWKAGYEYLCIDDCWAPORSEGRLOADPQRFPHGIROLANYVHSGKLGIYAD 136
Db 318 AADTWQ-----IBPDLWQTVEPAPPKKEDPVILLEHRCRDAAAKRIARELR---FSK 367
Qy 137 VGNKTCAGFGSGYDIDAQTFADMGVDLLKFKGCGYCDLSLENLADGKYKMSLALNR 196
Db 368 MVNE-IEGWPHA-----EFDG-FCDLNCLTQAEKY-----GK 398
Qy 197 SIVVSCSEMPLY 207
Db 399 KHYFPCSLPLF 409

RESULT 7
US-11-038-284-36
; Sequence 36, Application US/11038284
; Publication No. US20050246793A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
; APPLICANT: DEBET, MARTINE
; APPLICANT: GIDLEY, MICHAEL, JOHN
; APPLICANT: JOBLING, STEPHEN, ALAN
; APPLICANT: SAFORD, RICHARD
; APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL
; APPLICANT: WESTCOTT, ROGER, JOHN
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION
; FILE REFERENCE: 054163-5003-US
; CURRENT APPLICATION NUMBER: US/11/038,284
; PRIOR FILING DATE: 2005-01-21
; PRIOR APPLICATION NUMBER: US/10/056,454
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: PCT/GB96/01075
; PRIOR FILING DATE: 1996-05-03
; PRIOR APPLICATION NUMBER: GB 9607409.1
; PRIOR FILING DATE: 1996-04-10
; PRIOR APPLICATION NUMBER: GB 9509229.2
; PRIOR FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Lathyrus sp.
US-11-038-284-36

Query Match 3.5%; Score 80; DB 7; Length 861;
Best Local Similarity 25.6%; Pred. No. 4.8;
Matches 32; Conservative 17; Mismatches 42; Indels 34; Gaps 6;

Qy 108 ADPQRF--PHGIROLANYVHSGKLK---GIYADVGNKTCAGF-----PGSFG 150
Db 395 APSRFGTPELKSILDRHELGLLVLMDIYVHSSNNVTLGLNMFDTGDGHYFPGSRG 454
Qy 151 YYDI-DAQTF-----ADMGVDLLKFKGCGYCDLSLENLADGKYKMSLALNR 195
Db 455 YHMMWDSRLFNYSWEVLRYLLSNARWLDEYKDFGPRFDGVTSMW--YTHHGLQVSP 512
Qy 196 RSIVY 200
Db 513 NYSEY 517

RESULT 8
US-11-057-058-62
; Sequence 62, Application US/11057058
; Publication No. US2005024400A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan
; APPLICANT: Maza, John
; TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
; FILE REFERENCE: SYN-011
; CURRENT APPLICATION NUMBER: US/11/057,058
; CURRENT FILING DATE: 2005-02-10
; PRIOR APPLICATION NUMBER: US 60/543,812
; PRIOR FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.3
```

```
; SEQ ID NO 62
; LENGTH: 1827
; TYPE: PRT
; ORGANISM: Rabbit sp.
US-11-057-058-62

Query Match      3.4%; Score 79.5; DB 7; Length 1827;
Best Local Similarity 18.3%; Pred. No. 15;
Matches 42; Conservative 29; Mismatches 61; Indels 97; Gaps 10;

QY 143 AGPGSGFYDIDAQTFADMGV-----DILKFDGCGYCD----- 175
DQ 1360 RAAP-----DPERNSTAEWTRILDFYNNYMKEDGLWIDMNEPSSFVNGTTTVCRNT 1413
QY 176 -----SLENADGCKHSLALN-----RTGRSIVYSCWPLYMPPFQKPNYTEIR- 220
DQ 1414 ELNYPYPPELTKRTDGLHFRMCTMETHILSDGSSVLHYDVHNLVGHWSQAKPTYDALQK 1473
QY 221 -----QYCNHW--RNFADIDDSWKSISILDMWTSFNQERIVDVAGP- 259
DQ 1474 TTGKRGIVISRSTYTAGRWAGHWLGDYVARWDDNMDKSLIIGMWFSLFG-----ISYTGAD 1529
QY 260 --GGWDDPDMLV-----IGNF-----GLSNQOQVTQM 284
DQ 1530 ICGFFNDSEYHLCTRWTOLGAFYPPARNHNIQFTRRQDPVSNQTFVEM 1578

RESULT 9
US-10-467-657-8554
; Sequence 8554, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 8554
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8554

Query Match      3.4%; Score 79; DB 6; Length 322;
Best Local Similarity 20.7%; Pred. No. 1.6;
Matches 78; Conservative 38; Mismatches 101; Indels 160; Gaps 23;

QY 2 QLENPELHGCALARFLALVSWDIPGARALDNLGLARTPTWGLHWRFMCNLDQCEBPD 61
DQ 27 QCSUVELHWCBAQ-----FAGIARLHPFVKVHVMMKWRMK----- 64
QY 62 SCISKELFMEAMELMVSEGWKDAG-----YELCIDDCWMAPOQDSGRILQA--- 108
DQ 65 -----HLFQAEWTREMGRLKQALLQEAQDFVL-----DSGLIKSACF 102
QY 109 -----DPQRPFGIROLA---NYVHSKGLKL-----GIYADVGNKTCAGFPQSF 149
DQ 103 AKMAKSPYGLDKHSAREGVAALAYDKYVVPKGRNAVVRNRLDPAQV-----F 151
QY 150 GYVIDAQTADGWGDLKFDGCYCDSELENADGCKHSLALNRTGRSIVYSCWPLYM 209
DQ 152 GYAMPETQVF-----GLSVPE-----ASRLKNLAQPY---YAAHATSRD-----SKLWPMENW 197
QY 210 P--FQKPNYTEIRQYCN---HWRNFADIDDSWKSISILDMWTSFNQERIVDVAGPGWND 264
; SEQ ID NO 62
; LENGTH: 1827
; TYPE: PRT
; ORGANISM: Rabbit sp.
US-11-057-058-62

Query Match      3.4%; Score 79.5; DB 7; Length 1827;
Best Local Similarity 18.3%; Pred. No. 15;
Matches 42; Conservative 29; Mismatches 61; Indels 97; Gaps 10;

QY 143 AGPGSGFYDIDAQTFADMGV-----DILKFDGCGYCD----- 175
DQ 1360 RAAP-----DPERNSTAEWTRILDFYNNYMKEDGLWIDMNEPSSFVNGTTTVCRNT 1413
QY 176 -----SLENADGCKHSLALN-----RTGRSIVYSCWPLYMPPFQKPNYTEIR- 220
DQ 1414 ELNYPYPPELTKRTDGLHFRMCTMETHILSDGSSVLHYDVHNLVGHWSQAKPTYDALQK 1473
QY 221 -----QYCNHW--RNFADIDDSWKSISILDMWTSFNQERIVDVAGP- 259
DQ 1474 TTGKRGIVISRSTYTAGRWAGHWLGDYVARWDDNMDKSLIIGMWFSLFG-----ISYTGAD 1529
QY 260 --GGWDDPDMLV-----IGNF-----GLSNQOQVTQM 284
DQ 1530 ICGFFNDSEYHLCTRWTOLGAFYPPARNHNIQFTRRQDPVSNQTFVEM 1578

RESULT 9
US-10-467-657-8554
; Sequence 8554, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 8554
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8554

Query Match      3.4%; Score 79; DB 7; Length 368;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 50; Conservative 32; Mismatches 79; Indels 34; Gaps 11;

QY 176 SLENADGCKHMS--LALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADID 233
DQ 175 ALERLRDISKELSVPIIVKESGNGI--SMETAKLLYSYGKFNFTSGGGTNWIAEIMR 232
QY 234 D---SWK--SIKSLDMWTSFNQERIVDVAGPGWNDPDMVLVGNFGL--SWNQQTOMAL 286
DQ 233 DIRGNWKAESAQNFLDWGVTAASIEVR---YSIPDAFLVSGGIRSGLDAAKAL 288
QY 287 WAI---MAAPLFMS-----NDLRHISPOAKA---LLQDKDVIAINQDP---LGK-Q 327
DQ 289 GADIAGMALPVLKGAEGKESLEQFFRKIIIFELKATMMLTGSKNVEALKRSSIVILGKLK 348
QY 328 GYQLRQGNFEWFER 342
DQ 349 EWAERYGINLSIYEK 363

RESULT 11
US-10-510-386-236
; Sequence 236, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 236
; LENGTH: 622

; SEQ ID NO 62
; LENGTH: 1827
; TYPE: PRT
; ORGANISM: Rabbit sp.
US-11-057-058-62

Query Match      3.4%; Score 79.5; DB 7; Length 1827;
Best Local Similarity 18.3%; Pred. No. 15;
Matches 42; Conservative 29; Mismatches 61; Indels 97; Gaps 10;

QY 143 AGPGSGFYDIDAQTFADMGV-----DILKFDGCGYCD----- 175
DQ 1360 RAAP-----DPERNSTAEWTRILDFYNNYMKEDGLWIDMNEPSSFVNGTTTVCRNT 1413
QY 176 -----SLENADGCKHSLALN-----RTGRSIVYSCWPLYMPPFQKPNYTEIR- 220
DQ 1414 ELNYPYPPELTKRTDGLHFRMCTMETHILSDGSSVLHYDVHNLVGHWSQAKPTYDALQK 1473
QY 221 -----QYCNHW--RNFADIDDSWKSISILDMWTSFNQERIVDVAGP- 259
DQ 1474 TTGKRGIVISRSTYTAGRWAGHWLGDYVARWDDNMDKSLIIGMWFSLFG-----ISYTGAD 1529
QY 260 --GGWDDPDMLV-----IGNF-----GLSNQOQVTQM 284
DQ 1530 ICGFFNDSEYHLCTRWTOLGAFYPPARNHNIQFTRRQDPVSNQTFVEM 1578

RESULT 9
US-10-467-657-8554
; Sequence 8554, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 8554
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8554

Query Match      3.4%; Score 79; DB 7; Length 368;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 50; Conservative 32; Mismatches 79; Indels 34; Gaps 11;

QY 176 SLENADGCKHMS--LALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADID 233
DQ 175 ALERLRDISKELSVPIIVKESGNGI--SMETAKLLYSYGKFNFTSGGGTNWIAEIMR 232
QY 234 D---SWK--SIKSLDMWTSFNQERIVDVAGPGWNDPDMVLVGNFGL--SWNQQTOMAL 286
DQ 233 DIRGNWKAESAQNFLDWGVTAASIEVR---YSIPDAFLVSGGIRSGLDAAKAL 288
QY 287 WAI---MAAPLFMS-----NDLRHISPOAKA---LLQDKDVIAINQDP---LGK-Q 327
DQ 289 GADIAGMALPVLKGAEGKESLEQFFRKIIIFELKATMMLTGSKNVEALKRSSIVILGKLK 348
QY 328 GYQLRQGNFEWFER 342
DQ 349 EWAERYGINLSIYEK 363

RESULT 11
US-10-510-386-236
; Sequence 236, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 236
; LENGTH: 622
```

```
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-236

Query Match
Best Local Similarity 3.4%; Score 78.5; DB 6; Length 622;
Matches 65; Conservative 24; Mismatches 94; Indels 121; Gaps 16;

QY 97 APQRSEGRQLADPQRFPHGIRQLANYVHSGKLGIVADVGNKTCAGPFGSFG-----150
Db : : : : :
254 ADYRNAQGRILSGPE-----YLTIV-----FKGDTGAELTIV 284
QY 151 YDIDAQTFADWG-----VDLLKFDGVCYDLEN-----LADGY--KMSLALNRTGRSIV 199
Db : : : : :
285 NYEPARGNADWDGSGYGRVDRFLAGVAYLDGERPSFVMARGYYTRTVLVAYNFRGGKLT 344
QY 200 YSCEWPLYMWPQK--PNVTEIRQVCNHRNFADDDSKWSIKSILDWTSFNOERI-----253
Db : : : : :
345 -----KLWTFDSAPGNGAYAGQGNHSLSVADVGDGK-----DELIYCAM 385
QY 254 -VDVAGPG-----GWNDDPMLVIGN-----FGLSMNQOVTOMALWA 288
Db : : : : :
386 AVDHGKGLYSTGWGHDAMHTGNDLPSRPGLEVQVHENSNSPYGLSFRDAKTGKIWG 445
QY 289 I-----MAAPLFMSNDRLHISPOAKA-----LLQDKDVTAINDPDLGKQGYQLRQGDN 336
Db : : : : :
446 VHAGKDVGRGMAADI-----DPRYEGAENVANGSLYTAKGV-----KIGNTLPSSTN 492
QY 337 FEVW 340
Db : : : : :
493 FGIW 496

RESULT 12
US-10-821-234-1141
; Sequence 1141, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt seq_genes Version 1.0
; SEQ ID NO 1141
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1141

Query Match
Best Local Similarity 3.4%; Score 78; DB 6; Length 760;
Matches 40; Conservative 25; Mismatches 59; Indels 84; Gaps 8;

QY 49 RPMCNDLCOEPPDSCISEKLFMEMAELMVSEGWDAGYEYLCIDCW-----96
Db RIVTVYFSAHDNDANLQMDIFLEEDGNVNSKLTGDSVMYHCWNEAWTRPDLVPVFGWG 421
QY 97 -----APQDSDSGRLQADP-----QRFPHGIRQLANYVHSGKLGIVADVGNKTCAGPFGS 148
Db : : : : :
422 QAVDSTPOENSQDMYRCGPASVQATKHG-----HVCFPQDAP 458
QY 149 FGYYDIDAQTFADWGVDLL-----KFDGVCY-----DSLENLADGY 184
Db : : : : :
459 F-----VFAEVNSDLIYITAKDKGTHVENVDATHGKLIIVTKQIGGDMIDITDY 510
QY 185 KHMS-----LALNRTGRSIVSCEWPL 206

; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-236

Query Match
Best Local Similarity 3.4%; Score 78; DB 6; Length 786;
Matches 76; Conservative 39; Mismatches 106; Indels 180; Gaps 20;

QY 5 NPETHLGCALALRFLALVSDIPGARALDNGLAR-----38
Db : : : : :
280 NGYLHGHAKAM-----FVDFGLAKERGCGCYLRVDDTNPEAEKEEYINHI 325
QY 39 TPTWGLHWEFMCNLDQEE-----PDSCISE--K 67
Db : : : : :
326 BEIVKMGWEPPKELYDLAVELIRRHAYVDHQTADRIKEYREKQNSPWRDRPIESLK 385
QY 68 LFMEMAELMVSEGWDAGYEYLCIDCWAPQDSEGRLOADPQRFPHGIRQLANYVHSG 127
Db : : : : :
386 LFDWMRGIIIEG-----KATLRMKQDMQSDNFMYDLIAY-----421
QY 128 GLKLGIVADVGNKTCAGPFGSFGYVDIDAQTFADWGVLDLLKPDGVCYDLENLADGYKHM 187
Db : : : : :
422 RIKFAPHPKAGDKWCI-YPS---YD-----YAHCTVDSLENIT---HS 457
QY 188 SLALN-RTGRSIVSCEWPL-----YMHFFQKPNVTEIRQVCNHRNFADIDDSMKS 238
Db : : : : :
458 LCTLEFETRASY--WLLHSLSLYMPYVWEYGRNLNT-----493
QY 239 IKSILDWTSFN---QERIVDVAGPGWNDPMLVIGNFGLSMNQOVTOMALWAIM-----290
Db : : : : :
494 -NTVNSRKLNYYIVTNKYVD-----GWDDPRLILS--GLR-RRGVTSTAINAFVRGIGI 544
QY 291 -----AAPLFMSNDRHI-----SPOKALLQDKDVIAIN 320
Db : : : : :
545 TRSDGSMIHVSRLHREELNKTAPRTMVVLNPLKVVITN 585

RESULT 14
US-10-131-826A-450
; Sequence 450, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
```



```

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 450
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-450

Query Match 3.3%; Score 77.5; DB 6; Length 522;
Best Local Similarity 23.3%; Pred. No. 4.4;
Matches 85; Conservative 48; Mismatches 107; Indels 125; Gaps 23;

Qy 102 SEGRLOADPQPFPHGIQI-----ANYVH-----SKGLKGIYADVGN---KTCA 143
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
145 SYNKLQALAPDLFHLGRKLTTLHMRANAIFQVPVRFQDCRSLK---FLDIGYNQLSLA 201
Qy 144 --GFPGSFGYDIDAQTFADMGVDLLKFDGCYCDLSLENLADGYKMSLALNRTGRSIVVS 201
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 RNSFAGLFKLTELHLE-----HNDLVKVNFAHFPRILSL-----HSLCLRKNKVAIVVS 250
Qy 202 C---EWPL-----YMWPP--FQKPNYTEIRQYCNHWRNFAD--IDDSWKSISIL 243
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 SLDVWNLEKMDLSGNEIEMEPHFVETVPHLQSLQDSNRITYTEPRILNSWKSLSLTSI- 309
Qy 244 DWTSPQERIVDVAGPGGWNDDMLVIGNFGLSWN--QOVTQMALWAIMAAPLFMSN--- 298
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
310 -----TLAGNL-----WDCGRNVCALASG-----LSNFQG 334
Qy 299 ----DLRHTSPQAKALLQDKDVI-AINQDPLGKQGVQLRQGDNFWEWERPLSGLAWAVAM 353
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
335 RYDGNLQCASPE---YAQGEDVLDVAVYFHLCEGDAE-----PTSGHLLS-AV 378
Qy 354 INROEIGGPRSYTIASVLSKGV---ACNPACFITQLLPVKRKLGFYEWTSRLSRHINPT 410
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 TNRSDLGPPASSATTILADGEGQHDGTPEPA---ITVALP-----GGEAENAVQIHKVT 430
Qy 411 GTVLL 415
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
431 GTMAL 435
```

```

RESULT 15
US-11-057-058-65
; Sequence 65, Application US/11057058
; Publication No. US20050244400A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan
; APPLICANT: Maga, John
; TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
; FILE REFERENCE: SYM-011
; CURRENT APPLICATION NUMBER: US/11/057,058
; CURRENT FILING DATE: 2005-02-10
; PRIOR APPLICATION NUMBER: US 60/543,812
; PRIOR FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 65
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Spinacia oleracea
US-11-057-058-65
```

```

Query Match 3.3%; Score 77.5; DB 7; Length 903;
Best Local Similarity 17.7%; Pred. No. 9.1;
Matches 84; Conservative 47; Mismatches 136; Indels 207; Gaps 22;

Qy 73 AELVSEGWKDAYEYLICDDCWMAQPDSEGRLOADPQRFH-GHIRQLANVHSGKLKL 131
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
343 AKIPLFVNMWTDIDY-----MDAYKDTLDVNFPLDKMKKFNVLNHNKNGOKY 389
Qy 132 GIYADVGV---NKTGAGPGSGFYDIDAQTFADMGVDLLKFDGCYCDLSLENLADGYKMS 188
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
390 VVILDPGISTNKT-----YETIYIRGMKH-D 413
Qy 189 LALNRTGRSIVSCEW--PLYMWPFOKPN-----YTEIRQYCNHWRNFADIDDSWKSIXS 241
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
414 VFLKRNKGPYLGVS-WPGPVYFPDFLKPALSALTFTWDEIKREL----NLLPVDGLMIDMNE 468
Qy 242 ILDWTIS-----FNQE 251
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 ISNFISSPIPGSTLDNPPYKINNSGVNLPINIKTIPPTAMHYGDIPEYVNVHNLFGYDEA 528
Qy 252 RIVDVA-----GPG-----GWND-----PDMVLIGNFGLS 276
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
529 RVTRAALIKLTEKRPFFVLSRSTFSGSGKYTAHWTGDNAAWTNDLVYSTPSMLDFGLFGI- 587
Qy 277 WNQOVTQMALWAIMAAPL--FMSNDLRHISPOAKALLQDKDVIQINQD--PIGKQGYQLR 332
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
588 -----PMVGADICGFLGNTTBEL---CRRWIQLGAFYFPFSDHSSLGTYQELY 633
Qy 333 QGDNFEVWERPLSGLAWAV-----AMINRQEIIGGPRSY-----TIAVAS---L 372
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
634 RWESVAASARKVLGRLRYTLFPYFTLYEAOQLNGIPIARPLFFSFPPDIKTYGISSQFLL 693
Qy 373 KGKGVACNP-----ACFITQLLP-----VKEKLGFIYEWTSRLSRHIN 408
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
694 KGKGVVSPVLKPGVSVTAYFPRGNWFDLFYTRSVTASTGRYVTLSPDPDHIN 747
```

Search completed: December 24, 2005, 09:07:58
Job time : 14 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 13:48:48 ; Search time 282 Seconds
(without alignments)
774.752 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MOLRRPELHLCALALRFLA.....RLRSHNPTGTVLQLENTM 421

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4168288 seqs, 259477437 residues

Total number of hits satisfying chosen parameters: 8336576

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO spool_p/US10602219/runat_23122005_151144_17285/app_query.fasta_1.583
-DB=Published Applications NA New -QFMT=fastcap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10602219@cgn 1.183 @runat_23122005_151144_17285
-NCPU=6 -ICPU=3 -NO MMAP -LARGEOVERLY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEVTIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New.*

1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US12_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US13_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US14_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US15_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	4.3	1515	7 US-11-073-112-5	Sequence 5, Appli
2	95.5	4.1	268685	6 US-10-933-025-22	Sequence 22, Appl
3	94	4.1	1241	6 US-10-750-185-56907	Sequence 56907, A
4	93	4.0	1082144	7 US-11-117-187-211	Sequence 211, App
5	91.5	4.0	2177	7 US-11-073-112-4	Sequence 4, Appli
6	90.5	3.9	2716	6 US-10-750-185-31179	Sequence 31179, A
7	89.5	3.9	1766	6 US-10-750-185-35182	Sequence 35182, A
8	88	3.8	98345	7 US-11-112-908-36	Sequence 36, Appli

9	88	3.8	127340	7 US-11-112-908-35	Sequence 35, Appl
10	87.5	3.8	2576	9 US-11-038-284-16	Sequence 16, Appl
11	87.5	3.8	2578	9 US-11-038-284-19	Sequence 19, Appl
12	87.5	3.8	5301	7 US-11-108-528-83	Sequence 83, Appl
13	86.5	3.7	1615	6 US-10-750-185-56545	Sequence 56545, A
14	85	3.7	1212	6 US-10-750-185-55673	Sequence 55673, A
15	85	3.7	1314	9 US-11-073-626-4	Sequence 4, Appli
16	84.5	3.7	3033	9 US-11-038-284-14	Sequence 14, Appl
17	84.5	3.7	3033	9 US-11-038-284-41	Sequence 41, Appl
18	84	3.6	2866	6 US-10-510-386-235	Sequence 235, App
19	83.5	3.6	1948	7 US-11-055-822-1	Sequence 1, Appli
20	83.5	3.6	1262	6 US-10-750-185-45064	Sequence 45064, A
21	82	3.5	1314	9 US-11-073-626-2	Sequence 2, Appli
22	82	3.5	2626	6 US-10-750-185-58314	Sequence 58314, A
23	81.5	3.5	889	6 US-10-750-185-37410	Sequence 37410, A
24	81.5	3.5	1335	6 US-10-467-657-8553	Sequence 8553, Ap
25	81.5	3.5	1335	6 US-10-374-954-4	Sequence 4, Appli
26	81.5	3.5	1548	6 US-10-630-203-5	Sequence 5, Appli
27	81.5	3.5	1548	6 US-11-103-037-6	Sequence 6, Appli
28	81.5	3.5	2380	7 US-11-103-240-161	Sequence 161, App
29	81.5	3.5	2399	6 US-10-750-185-35565	Sequence 35565, A
30	81.5	3.5	2498	6 US-10-750-185-45595	Sequence 45595, A
31	81.5	3.5	189252	7 US-11-121-086-54	Sequence 54, Appl
32	81	3.5	930	7 US-11-074-176-107	Sequence 107, App
33	81	3.5	4860	6 US-10-971-982-1	Sequence 1, Appli
34	81	3.5	168516	7 US-11-121-086-3	Sequence 3, Appli
35	80.5	3.5	1242	6 US-10-467-657-31	Sequence 31, Appl
36	80.5	3.5	1242	6 US-10-467-657-8371	Sequence 8371, Ap
37	80.5	3.5	6423	6 US-10-750-185-61474	Sequence 61474, A
38	80	3.5	1559	6 US-10-864-758-1	Sequence 1, Appli
39	80	3.5	1611	6 US-10-864-758-6	Sequence 6, Appli
40	80	3.5	5745	6 US-10-750-185-54380	Sequence 54380, A
41	80	3.5	35344	6 US-10-995-561-13307	Sequence 13307, A
42	80	3.5	60844	6 US-10-995-561-13359	Sequence 13359, A
43	79.5	3.4	867	6 US-10-454-437-363	Sequence 363, App
44	79.5	3.4	960	6 US-10-467-657-3253	Sequence 3253, App
45	79.5	3.4	1251	9 US-11-082-389-121	Sequence 121, App

ALIGNMENTS

RESULT 1
US-11-073-112-5
; Sequence 5, Application US/11073112
; Publication No. US20050260627A1
; GENERAL INFORMATION:
; APPLICANT: Hintz et al.
; TITLE OF INVENTION: Mannosidases and Methods for using the Same
; FILE REFERENCE: 62447-02
; CURRENT APPLICATION NUMBER: US/11/073,112
; CURRENT FILING DATE: 2005-03-03
; PRIOR APPLICATION NUMBER: US 10/089,211
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: PCT/US00/27210
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/157,341
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5
; LENGTH: 1515
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-11-073-112-5

Alignment Scores:
Pred. No.: 0.492 Length: 1515
Score: 99.00 Matches: 62
Percent Similarity: 35.74% Conservative: 32
Best Local Similarity: 23.57% Mismatches: 74
Query Match: 4.28% Indels: 95
DB: 7 Gaps: 18


```

; PRIOR APPLICATION NUMBER: 60/157,341
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 4
; LENGTH: 2177
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-11-073-112-4

```

Alignment Scores:		
Pred. No.:	6.48	Length:
Score:	91.50	Matches:
Percent Similarity:	41.72%	Conservative:
Best Local Similarity:	27.61%	Mismatches:
Query Match:	3.95%	Indels:
DB:	7	Gaps:
		11
		2177

US-10-602-219-12 (1-421) x US-11-073-112-4 (1-2177)

Qy	27	ProGlyAlaArgAlaLeuAspAsnGlyLeuAlaAargThrProThrMetGlyTrpLeuHis	46
Db	1059	CCCCGGACCCAGCGCAGTA-----CGAACCGTTCCTCGATTGAT	1097
Qy	47	TripGluaArgPheMet---CysAsnLeuAspCysGlnGluCluProAspSerCysIleSer	65
Db	1098	TTTTTTTGGGAAGCGCAGTCAATATTCCGACGCCAAGCTCGCCA-----TGGTCA	1142
Qy	66	GluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys-AspAlaG1	85
Db	1143	:::::CATCAGCTG-----GAATGGTGCACGACTCGTA	1172
Qy	85	yTryGluTyTrLeu-----CysIleAspAs	93
Db	1173	CTACGAGTAGCTGATCAAGATGATACGCTACGATCCGAAACGGCTTTGGCCTCTACCGGA	1232
Qy	93	pCystTrpMetAlaProGlnAraAspSerGluGlyAraGLeuGlnAlaAspProGlnAraPh	113
Db	1233	CCGCTGGCTCGCAGCTGCCGAGTCGAGCATCAACCATCTGGCTTGCACCCGCCACCCG	1292
Qy	113	eProHisGlyIleAargGlnLeuAlaAsnTyValHisSerLysGlyLeuLysLeuGlyI1	133
Db	1293	CCCA---GACGTGACTTCTTTGGCCACTTAC-----AACGAGGAGCATCAGCTGGCGCCT	1343
Qy	133	eTryAlaAspValGlyAsnLysThrCysAlaGlyPheProGlySerPhe-----GlyTy	151
Db	1344	GACCAGCCAA-----CACCTGACCTGCTTCGAC---GGTGGHAGCTTTCTGCTTGGTGG	1394
Qy	151	rTryAspIleAspAlaGlnThrPheAlaAspTripGlyValAspLeuLeuLysPheAspG1	171
Db	1395	GACATTGCTGGACCGCCAGGACTTTGTGCAGATCTGGCCTTGACCTTGTCGCC-----GG	1448
Qy	171	yCysTyr	173
Db	1449	CTGCCAC	1455

RESULT 6

```

US-10-750-185-31179
; Sequence 31179, Application US/10750185
; Publication No. US2005026063A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DeNISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR I
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,
; PRIOR FILING DATE: 2003-12-31
; CURRENT APPLICATION NUMBER: US 60/437,48

```

```

; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PATENTIN version 3.1
; SEQ ID NO 31179
; LENGTH: 2716
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-31179
1986680843346

```

Alignment Scores:		
Pred. No.:	11.8	Length: 2716
Score:	90.50	Matches: 48
Percent Similarity:	30.84%	Conservative: 22
Best Local Similarity:	21.15%	Mismatches: 82
Query Match:	3.91%	Indels: 76
DB:	6	Gaps: 11

US-10-602-219-12 (1-421) x US-10-750-185-31179 (1-2716)

Qy	38	ArgThrProThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCys	--- 56
Db	1674	AGGTTCCAGCCTCGGGACCCCTCAGCTGGTAAAGAATC-----CACCTGCAATGCGGG	1727
Qy	57	GlnGluGluProAspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMet	76
Db	1728	AGACCTGAGTTCATCCCTGCGTTGGGAAG-----	1757
Qy	77	ValSerGluGlyTrpLysAspAlaGlyTyrrGluTyrrLeuCysIleAspAspCysTrpMet	96
Db	1758	-----ATCCCCTGGAGAGGGGAAGGCTCTCCACTCTGGCCCTGGAGAAATTTTCATGGACT	1811
Qy	97	AlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGly	116
Db	1812	GTATAGTCCATAGGTCGCAGAGTCGGACACAA-----CCAAAGCGACTTT-----	1856
Qy	117	IleArgGlnLeuAlaAsnTyrrValHisSerLysGlyLeuLysLeuGlyIleTyrrAlaAsp	136
Db	1857	-----CACTTCATTCCGGGCAAGGGGCTGGTATTGGCGGTT-----	1892
Qy	137	ValGlyAsnLysThrCysAlaGlyPheProGlySerPheGlyTyrrTyrrAspIleAspAla	156
Db	1893	-----TGTTACCAAGTTCCTCTCTTGGCTGGGGCAAGGCGGATGCT	1937
Qy	157	GlnThrPheAlaAspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrrCysAspSer	176
Db	1938	ACCTCTGTTGGCTC-TGG-----TGCAAGCTGT	1963
Qy	177	LeuGluAsnLeuAlaAspGlyTyrrLysHisMetSerLeuAlaLeuAsnArgThrGlyArg	196
Db	1964	GTGCAGAGTGCTGCTCATGGA-----GGTGCTCAGGTGGGCAGGTGAGGACTG	2011
Qy	197	SerIleValTyrrSerCys-----GluTrpProLeu	206
Db	2012	GGAGTGCCAGGGGCTGTGGCCAGGGCCCTGACAGCCTCTGCCACTCTACTGGACACAG	2071
Qy	207	TyrMetTrpProPheGlnLys-----ProAsnTyrr	216
Db	2072	GCATGGTGGCGCAGCAGTCCCTAGGAGTGGCTCTCTGCCAAGCCCTCTTGTTCACCGCTC	2131
Qy	217	ThrGluIleArgGlnTyrr-----CysAsnHisTrpArgAsnPhe	229
Db	2132	ACGTCTCTCCAGGCCCAAGGAGTGCACCAACCCCTCTCTGCTGCCCTGGGGAGACCCCT	2191
Qy	230	AlaAspIleAspSerTrp	236
Db	2192	TCCCACTGGGACCAGTCATGG	2212

RESULT 7
 US-10-750-185-35182
 ; Sequence 35182, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.

; APPLICANT: DeNISE, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM1100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 35182
 ; LENGTH: 1766
 ; TYPE: DNA
 ; ORGANISM: Bovine 19866880928175
 US-10-750-185-35182

Alignment Scores:
 Pred. No.: 8-23 Length: 1766
 Score: 89.50 Matches: 57
 Percent Similarity: 31.33% Conservative: 21
 Best Local Similarity: 22.89% Mismatches: 85
 Query Match: 3.87% Indels: 85
 DB: 6 Gaps: 15

US-10-602-219-12 (1-421) x US-10-750-185-35182 (1-1766)

Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAla----- 37
 Db 481 CTCTGTCATGGATTTCCAGGCAAG-----AATCTGGATCATGTTGCCATTTC 531
 Qy 38 -----ArgThrProThrMetGly-----TrpLeuHisTrpGlu-Ar 49
 Db 532 CTCTCCAGGAGATCTTGCCAAACCCCAAGGATCGAACCCATTTCCTCTCACTGGGAGGG 591
 Qy 49 gPheMetCysAsnLeuAspCysGlnGluProAspSerCysIleSerGluLysLeuPh 69
 Db 592 GATTCCTTCCACTAGCCAGCCAGGGAAGCCCCCACTAGTGTGTAGCACATTTTATA 651
 Qy 69 eMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAsp-----AlaG1 85
 Db 652 CCTTCAA-----CAGAAAGGTGGGCGAGAAACAGTCTTGGGGGCGAG 693
 Qy 85 YTrpGluTrpLeu---CysIleAspAsp-----CysTrpMetAlaProGlnArgAspSe 102
 Db 694 ATATGCCCTCTGCTCTCTGTGTTCTTCTCTGTTCTCTGGAGGCTGGACATCAGG----- 748
 Qy 102 rGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAs 122
 Db 749 -GAAGGGATGCTGGTGCCTCT----- 769
 Qy 122 nTyrrValHisSerLysGlyLeuLysLeuGlyIleTyrrAlaAspValGlyAsnLysThrCy 142
 Db 770 -----GATGTCAGAGACCAAGCTTGGCCACTTAAGTCAGCGGGAAGTGCAGACTG 822
 Qy 142 sAlaGlyPheProGlySerPheGlyTyrrTyrrAspIleAspAlaGlnThrPheAlaAspTr 162
 Db 823 CCAGGGGCTGCACAGATCATCGGA-----GGCTG 852
 Qy 162 pGlyValAspLeuLeuLys-----PheAspGlyCysTyrrCysAs 175
 Db 853 GAGAACTGGGCTTACAAAGAACAGCAGGAGCAATTTCTGAGGCTGGGATGCAATATGGGCAGA 912
 Qy 175 pSerLeu-----GluAsnLeuAlaAspG1 183
 Db 913 TGCCTCTAGGGGAAGCAGGCTGGGTAGTGGGTGACTTCTTAGGAAGAGGCGCACTGG 972
 Qy 183 YTrpLysHisMetSerLeuAlaLeuAsnArgThr-----GlyArgSerIleVa 199
 Db 973 CTATGTCAGTGTGCCATACCCGACCCAGGATTTTAGTTTCCAGGGGGAGAGATCTCACT 1032

Qy 199 lTyrrSerCys---GluTrpProLeuTrpMetTrpProPheGlnLysProAsnTyrrThrG1 218
 Db 1033 TGCCTTCTGTGTCTCTGGGTGTC-----CCTTCGTCTACCA 1071
 Qy 218 uIleArgGlnTyrrCysAsnHisTrp 226
 Db 1072 GACTGAACCTGTAGGAGGACACTGG 1096

RESULT 8

US-11-112-908-36
 ; Sequence 36, Application US/11112908
 ; Publication No. US20050260659A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Harris, Cole
 ; APPLICANT: Davis, Lisa M.
 ; TITLE OF INVENTION: Breast
 ; FILE REFERENCE: 04-164-US
 ; CURRENT APPLICATION NUMBER: US/11/112,908
 ; CURRENT FILING DATE: 2005-04-22
 ; PRIOR APPLICATION NUMBER: US 60/564,758
 ; PRIOR FILING DATE: 2004-04-23
 ; PRIOR APPLICATION NUMBER: US 60/575,978
 ; PRIOR FILING DATE: 2004-06-01
 ; PRIOR APPLICATION NUMBER: US 60/631,702
 ; PRIOR FILING DATE: 2004-11-30
 ; PRIOR APPLICATION NUMBER: US 60/633,826
 ; PRIOR FILING DATE: 2004-12-07
 ; NUMBER OF SEQ ID NOS: 511
 ; SOFTWARE: PatentIn version 3.3
 ; SEQ ID NO 36
 ; LENGTH: 98345
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (49350)..(49350)
 ; OTHER INFORMATION: n is a, c, g, or t
 US-11-112-908-36

Alignment Scores:
 Pred. No.: 4.55e+03 Length: 98345
 Score: 88.00 Matches: 49
 Percent Similarity: 32.4% Conservative: 15
 Best Local Similarity: 24.8% Mismatches: 75
 Query Match: 3.80% Indels: 58
 DB: 7 Gaps: 11

US-10-602-219-12 (1-421) x US-11-112-908-36 (1-98345)

Qy 225 HisTrpArgAsnPheAlaAspIleAspSerTrpLysSerIleLysSerIleLeuAsp 244
 Db 76964 CATTCGCGA----- 76972
 Qy 245 TrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrpAsnAsp 264
 Db 76973 -----CAGAGCGGGCAGCGCCCGCTCGGGCCGAGAGG----- 77008
 Qy 265 ProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsn-----GlnGlnValThr 282
 Db 77009 ---GAGCGGTACAGCGGGGGCTGGCACCGCGGTGGAACTCGCGCGGGGTCC 77065
 Qy 283 GlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHis 302
 Db 77066 GCCCTGGCCCTCTGGCCGCCGCCCGCTCGCTTGTCTTCG----- 77110
 Qy 303 IleSerProGlnAlaLysAlaLeuLeuGlnAspValIleAlaIleAsnGlnAsp 322
 Db 77111 CTCGCCCTCTCCCCCGCCCGCTCGCTTCACAGCGCGCGAGAGCCTAACCGCGGT 77170
 Qy 323 ProLeuGlyLysGlnGlyTyrrGlnLeuArgGlnGlyAspAsnPheGluValTrpGluArg 342
 Db 77171 CCTCTAGGAACCTCGGGCGGGCAGCACCCCGGGGA-----TTCTGTGGGTCTCT 77221


```

Db 1258 GATGACTGAACATGTTTACGGCACCAGCATAGTTGTTACTTCTACTCTGGAGCTCGTGGT 1317
Qy 151 TyrTyAspIle---AspAlaGlnThrPhe----- 159
Db 1318 TATCATTTGATGGGATTCGCCGCTTTTAACTATGGAACTGGAGGACTATTAGGTAT 1377
Qy 160 -----AlaAspTTPGlyValAspLeuLeuLeuPheAspGlyCysTyrCysAsp 175
Db 1378 CTTCTCTCAAAATGCGAGATGGTGGTGGATGAGTTCAAAATTTGATGGATTTAGATTGAT 1437
Qy 176 SerLeuGluAsnLeuAlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGly 195
Db 1438 GGTGTGACATCAATGATG-----TATACTCACCGGATTTATCGTGGGATTCACCTGG 1491
Qy 196 Arg-----AlaAspTTPGlyValAspLeuLeuLeuPheAspGlyCysTyrCysAsp 203
Db 1492 AACTACGAGGAATPACTTTGGACTCGCAACTGATGTGGATGCTGTTGTGTATCTGATGCTG 1551
Qy 203 uTrpProLeu-TyrMetTTPProPheGlnLysProAsnTyrThrGluIleArgGlnTyrC 223
Db 1552 GTCAACGATCTTATTCATGGCTTTTCCAGATGCAATTAC----- 1592
Qy 223 YeAsnHisTTP-----ArgAsnPheAlaAspIle 232
Db 1593 -----CATTTGGTGAAGATGTTAGCGGAAT---GCCGACATT 1625

RESULT 12
US-11-038-284-19
; Sequence 19, Application US/11038284
; Publication No. US20050246793A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
; APPLICANT: DEBET, MARTINE
; APPLICANT: GIDLEY, MICHAEL, JOHN
; APPLICANT: JOBLING, STEPHEN, ALAN
; APPLICANT: SAFFORD, RICHARD
; APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL
; APPLICANT: WESTCOTT, ROGER, JOHN
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION
; FILE REFERENCE: 054163-5003-US
; CURRENT APPLICATION NUMBER: US/11/038,284
; CURRENT FILING DATE: 2005-01-21
; PRIOR APPLICATION NUMBER: US/10/056,454
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: PCT/GB96/01075
; PRIOR FILING DATE: 1996-05-03
; PRIOR APPLICATION NUMBER: GB 9607409.1
; PRIOR FILING DATE: 1996-04-10
; PRIOR APPLICATION NUMBER: GB 9509229.2
; PRIOR FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 19
; LENGTH: 2578
; TYPE: DNA
; ORGANISM: Solanum tuberosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (24)..(2564)
US-11-038-284-19

Alignment Scores:
Pred. No.: 24.8 Length: 2578
Score: 87.50 Matches: 43
Percent Similarity: 37.14% Conservative: 22
Best Local Similarity: 24.57% Mismatches: 49
Query Match: 3.78% Indels: 61
DB: 9 Gaps: 9

US-10-602-219-12 (1-421) x US-11-038-284-19 (1-2578)
Qy 108 AlaAspProGlnArgPhe-----ProHisGlyIleArgGlnLeuAlaAsnTyrValHis 125
||| ||||| ||| : : : : : |||

```

```

Db 1140 GCACCAACGACGCGCTTTTGGAAACGCCGACGACCTTAAGTCTTTGATTGATAAAGCTCAT 1199
Qy 126 SerLysGlyLeuLysLeu----- 131
Db 1200 GAGCTAGGAATTTGTTTCTCATGGACATTTGTTTCACGCCATGCATCAATAATACTTTTA 1259
Qy 132 ---GlyIleTyrAlaAspValGlyAsnLysThrCysAlaGlyPheProGlySerPheGly 150
Db 1260 GATGACTGAACATGTTTTCACGGCACCAGCATAGTTGTTTCACTCTCGGAGCTCGTGGT 1319
Qy 151 TyrTyAspIle---AspAlaGlnThrPhe----- 159
Db 1320 TATCATTTGATGGGATTCGCCGCTTTTAACTATGGAACTGGAGGACTATTAGGTAT 1379
Qy 160 -----AlaAspTTPGlyValAspLeuLeuLeuPheAspGlyCysTyrCysAsp 175
Db 1380 CTTCTCTCAAAATGCGAGATGGTGGTGGATGAGTTCAAAATTTGATGGATTTAGATTGAT 1439
Qy 176 SerLeuGluAsnLeuAlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGly 195
Db 1440 GGTGTGACATCAATGATG-----TATACTCACCGGATTTATCGTGGGATTCACCTGG 1493
Qy 196 Arg-----SerIleValTyrSer-CysG 203
Db 1494 AACTACGAGGAATPACTTTGGACTCGCAACTGATGTGGATGCTGTTGTGTATCTGATGCTG 1553
Qy 203 uTrpProLeu-TyrMetTTPProPheGlnLysProAsnTyrThrGluIleArgGlnTyrC 223
Db 1554 GTCAACGATCTTATTCATGGCTTTTCCAGATGCAATTAC----- 1594
Qy 223 YeAsnHisTTP-----ArgAsnPheAlaAspIle 232
Db 1595 -----CATTTGGTGAAGATGTTAGCGGAAT---GCCGACATT 1627

RESULT 12
US-11-108-528-83
; Sequence 83, Application US/11108528
; Publication No. US20050261189A1
; GENERAL INFORMATION:
; APPLICANT: Larsen, Glenn
; APPLICANT: Marvin, Martha
; APPLICANT: Li, Dean Y.
; APPLICANT: Wang, Elizabeth
; APPLICANT: Chen, C. M. Amy
; APPLICANT: Shah, Steven M.
; TITLE OF INVENTION: METHODS OF PROMOTING CARDIAC CELL
; FILE REFERENCE: HYDR-P01-041
; CURRENT APPLICATION NUMBER: US/11/108,528
; CURRENT FILING DATE: 2005-04-18
; PRIOR APPLICATION NUMBER: US 60/563,137
; PRIOR FILING DATE: 2004-04-16
; PRIOR APPLICATION NUMBER: US 60/598,368
; PRIOR FILING DATE: 2004-08-02
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 5301
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-108-528-83

Alignment Scores:
Pred. No.: 71.4 Length: 5301
Score: 87.50 Matches: 96
Percent Similarity: 32.97% Conservative: 55
Best Local Similarity: 20.96% Mismatches: 157
Query Match: 3.78% Indels: 152
DB: 7 Gaps: 22

US-10-602-219-12 (1-421) x US-11-108-528-83 (1-5301)
Qy 46 HisTTP---GluArgPheMetCys---AsnLeuAspCysGlnGluProAspSerCys 63

```



```
QY 92 AspAspCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspPro--- 110
Db 1153 GATGATTTATTTATGAGCGATTGAGTATGAGGATACGAGGCTAAGGTGCTCCATTGA 1094
QY 111 GlnArgPheProHisGlyIleArgGlnLeuAlaAsnTrpValHisSerLys----- 127
Db 1093 GTGAGAGCGCACAGGAGGAGCTTGCTCTGCTTTGA-CATTTCAAGAGGCGACTC 1035
QY 127 ----- 127
Db 1034 CTACGAGGAGCTGCTTTCTTTTAGAAGTTTACCTTTAATGTTCTATTATCATGGCCT 975
QY 128 -----GlyLeuLysLeuGlyIleTrp 134
Db 974 TGTCTGAGCAGCTGGAGAGGTAGAGTAGCGGTGCGGGGGCTTTGGGGGGAGACAG 915
QY 135 AlaAspValGlyAsnLysThrCysAla---GlyPheProGlySerPheGlyTrpTrpAsp 153
Db 914 GCGGGGTCTGTCACCTACGTAGCCAGCTTGGGGCGCGCGGAGCGCTGGGCACGGGGCG 855
QY 154 IleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLeuLysPheAspGlyCysTrp 173
Db 854 AGAGCTGTGTCACG-----GACTGG----- 834
QY 174 CysAspSerLeuGlnAsnLeuAlaAspGlyTrpLysHisMet---SerLeuAlaLeuAsn 192
Db 833 -----CAGCCCGCGCGCCAGCAGCGCCACCTCCGCGGCTTGAGCTGCAGG 786
QY 193 ArgThrGlyArgSerIleValTrpSerCysGluTrpProLeuTrpMetTrpProPheGln 212
Db 785 CGCGCGGGCAGCAGCT-GTTTACTCTGTGTCGTCTGCTCTGTTTCTGTCGCG----- 733
QY 213 LysProAsnTrpTrpGluIleArgGlnTrpCysAsnHisTrpArgAsnPheAlaAspIle 232
Db 732 ---CCGGCTCCACCGCAGCAGCAGCTTTCACATCTCTGG----- 694
QY 233 AspAspSerTrpLysSerIleLysSerIleLeuAspTrpThrSerPheAsnGlnGluArg 252
Db 693 -----CTCTGGCCTGGCCAGGCGCTTTTCTCACTATGAGCTCTGGGGCTGTGGACAGA 640
QY 253 IleValAspValAlaGlyProGlyGlyTrp-----AsnAspProAspMetLe 268
Db 639 ATCCCAACTTGAG---CCCGGGCTGGATTGTTGCTTCGCTCTGTCGACTGACTCAT 583
QY 268 uValIleGlyAsnPheGlyLeuSerTrpAsnGln---GlnValThrGlnMetAlaLeuTr 287
Db 582 AGGACCCGCGAGCCTGGGCTCAGTTTCCCATGTGTCCATATGTCACATAATCAGATTGATTTG 523
QY 287 palalie-MetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSerProGlnA 307
Db 522 GCAAAATGTTTGGCAGTAGAGAGATTCTTCAAAATGATGTCCAGTGTCCAGGCCCAAGAT 463
QY 307 lalysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspProLeuGlyLysG 327
Db 462 ACAAGCA-----GATAGCTCAGGTGGCC 439
QY 327 lNGLyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrpGluArgProLeuSer---G 346
Db 438 CTGGTCACACCTGGCTCAAGGG-----CCTCCTCTCTGGCCTTAC 397
QY 346 lylLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGlyLys-----ProA 363
Db 396 TGCTGCCCTGGAAAGCTTCCTC-----CAGTTTGGGGGCTGGGAGTCCCCCT 352
QY 363 rgsrTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnProAlaCysP 383
Db 351 GTGGTCCCCCCTAGCAGCAGTGTGCGGAGGAGGCCACTGGCTGTGTGCTGTGTGTG 292
QY 383 heIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThrSer 401
Db 291 TGCTGACCCAGAGGACACACACACTGCGCCAGGCGCTGGGT-----TGGAGCTCG 245
```

```
RESULT 14
US-10-750-185-55673
; Sequence 55673, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KEER, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 55673
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Bovine 19856880894405
US-10-750-185-55673

Alignment Scores:
Pred. No.: 16.2 Length: 1212
Score: 85.00 Matches: 41
Percent Similarity: 30.57% Conservative: 18
Best Local Similarity: 21.24% Mismatches: 62
Query Match: 3.67% Indels: 72
DB: Gaps: 8

US-10-602-219-12 (1-421) x US-10-750-185-55673 (1-1212)
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 16 GCCTGTCACTGCGCCCAAGGAGGTGGTTGCAGACTCC----- 54
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAla-----AspIle 232
Db 55 -----AGAAACACTGCACCTGGGGGCTCTGTGTCGGCCTCTCCTGTG 96
QY 233 AspAspSerTrpLysSerIleLysSerIleLeuAspTrpThrSerPheAsnGlnGluArg 252
Db 97 GACCAACATGTGTTGATTGGGGTCTCTCATTCCTGGGCGCTGTTCTTTGTATG--- 153
QY 253 lIleValAspValAlaGlyProGlyGlyTrpAsn-----AspProAspMetLeuVal 269
Db 154 -----TGGAACAAAGGGGTTGACCTTGATTGCAAGTA 186
QY 270 Ile----- 270
Db 187 TTAGAACCCTCAGTTCGGTTTAGTCTTTCTGAGGAACCTCAAGAGAGGCTTGTGTACA 246
QY 271 -----GlyAsnPheGlyLeuSerTrpAsnGlnValThrGlnMetAlaLeuTrp 287
Db 247 GCGAACCTCGGAGATGACAGGCCCCAGCTGG-----GTCGTG 282
QY 288 AlalleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSerProGlnAla 307
Db 283 TCTCAGGTTGTCTGCTCTGCTCGGCTCCCTGGGTGGAGTTTGTTCCTCGGCAGCC 342
QY 308 LysAlaLeuLeuGlnAspLys-----AspValIleAlalleAsn 320
Db 343 TCGCCCCGCTCAGGCGCAGGAGGCGCTCGGAGCTCTCGATCTGTGTGTGTACAGCT 402
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsn---PheGluVal 339
Db 403 GACAGCCCCCAAGAAAGGGGCCACCTCTCTCCATGGTTCTAGCCCTGGGAGCCCT 462
QY 340 TrpGluArgProLeuSerGlyLeuAlaTrpAlaValAla 352
```

Qy 173 -----TyrCysAspSerLeuGluAsnLeuAlaAspGlyTyr 184
Db 703 GTTTCAAAGGCTGGGTTTTCGCTCATATTCAACTCACACCCAAAGAGCGACCGTAC 762
Qy 185 LysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrSerCysGluTrp 204
Db 763 AAGATGTG-----CTGTGGTCTATGATGGTGAATAT 795
Qy 205 ProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArgGlnTyrCysAsn 224
Db 796 GGGTTC-----TTTTTGGCCCAACGAGTATGGGGTATGATCAAAAGTCTGTGAC 843
Qy 225 HisTrpArgAsnPheAla 230
Db 844 GAGTTCCCTGGTTCCTCT 861

Search completed: December 26, 2005, 16:10:01
Job time : 810 secs

Db 463 CCCTCTCTCTGCTCCAGCTCTGTGTGGGACCTGGCT 501
US-11-073-626-4
RESULT 15
; Sequence 4, Application US/11073626
; Publication No. US20050244926A1
; GENERAL INFORMATION:
; APPLICANT: KUROSAWA, KEIKO
; APPLICANT: HIROKAWA, KOZO
; APPLICANT: KAJIYAMA, NAOKI
; TITLE OF INVENTION: Novel fructoysl peptide oxidase
; FILE REFERENCE: 227590US0
; CURRENT APPLICATION NUMBER: US/11/073,626
; CURRENT FILING DATE: 2005-03-08
; PRIOR APPLICATION NUMBER: US/10/232,655
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Eupenicillium terrenum
US-11-073-626-4
Alignment Scores:
Pred. No.: 18.2 Length: 1314
Score: 85.00 Matches: 46
Percent Similarity: 33.19% Conservative: 29
Best Local Similarity: 20.35% Mismatches: 71
Query Match: 3.67% Indels: 80
DB: 9 Gaps: 12
US-10-602-219-12 (1-421) x US-11-073-626-4 (1-1314)
Qy 32 LeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHisTrpGluArgPheMet 51
Db 343 CTCGATCGGGCATTTGGGCTGGAGAAGACGAACGTTTGGCTG----- 384
Qy 52 CysAsnLeuAspCysGlnGluProAspSerCysIleSerGluLysLeuPheMetGlu 71
Db 385 -----GAATCTGAAGATGAGATCTCGCCAAAGCCCGAATTTACG 426
Qy 72 MetAlaGluLeuMetValSerGluGlyTrpLysAspAlaGlyTyrGluTyrLeuCysIle 91
Db 427 CGTGAACAAGTC-----AAGGGTGGAAAGGC-----TTATTTGCAC 465
Qy 92 AspAspCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGln 111
Db 466 GATGGAGGCTGGCTTGTCTGCACCAAG----- 492
Qy 112 ArgPheProHisGlyIleArgGlnLeuAlaAsnTyrValHisSerLysGlyLeuLysLeu 131
Db 493 -----GCTATCAATCGCATCGGAATTTTCTCCAGGACAAAGGTGTCAAGTTT 540
Qy 132 Gly-----IleTyrAlaAspValGlyAsnLysThr 141
Db 541 GCGTTTGGAGGTGCTGGACATTTTCAGCAACCTCTGTTGCGCGCTGATGGA---AAACT 597
Qy 142 CysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAlaAsp 161
Db 598 TGCATCGGACTTGAAACTACACACGA-----ACCAAGTACTTTGCTGAC 642
Qy 162 -----TrpGlyValAspLeuLysPhe---AspGlyCys 172
Db 643 AAGGTTGTCTTGGTGTGGTGGTGGTCCACCTTGGTGGATCTAGAAAGATCAGTGT 702

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 12:40:36 ; Search time 890 Seconds
(without alignments)
3911.695 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQLRNPHELGCALALRFLA.....RLRSHINPTGTVLLQLENTM 421

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 9793542 seqs, 4134689005 residues
Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO spool_p/US10602219/runat_23122005_151144_17253/app_query.fasta_1.583
-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10602219 @CGN_1_1_1549 @runat_23122005_151144_17253 -NCPUs=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	1266	3	US-09-993-059-11
2	2314	100.0	1266	5	US-10-103-327-11
3	2314	100.0	1266	7	US-10-602-219-11
4	2314	100.0	1266	7	US-10-602-220-11
5	2314	100.0	1266	8	US-10-851-388-11
6	2314	100.0	1266	9	US-10-984-389-11
7	2314	100.0	1278	3	US-09-993-059-7

8	2314	100.0	1278	5	US-10-103-327-7	Sequence 7, Appli
9	2314	100.0	1278	7	US-10-602-219-7	Sequence 7, Appli
10	2314	100.0	1278	7	US-10-602-220-7	Sequence 7, Appli
11	2314	100.0	1278	8	US-10-851-388-7	Sequence 7, Appli
12	2314	100.0	1278	9	US-10-984-389-7	Sequence 7, Appli
13	2314	100.0	1284	3	US-09-993-059-13	Sequence 13, Appl
14	2314	100.0	1284	5	US-10-103-327-13	Sequence 13, Appl
15	2314	100.0	1284	7	US-10-602-219-13	Sequence 13, Appl
16	2314	100.0	1284	7	US-10-602-220-13	Sequence 13, Appl
17	2314	100.0	1284	8	US-10-851-388-13	Sequence 13, Appl
18	2314	100.0	1284	9	US-10-984-389-13	Sequence 13, Appl
19	2314	100.0	1290	3	US-09-993-059-3	Sequence 3, Appli
20	2314	100.0	1290	5	US-10-103-327-3	Sequence 3, Appli
21	2314	100.0	1290	7	US-10-602-219-3	Sequence 3, Appli
22	2314	100.0	1290	7	US-10-602-220-3	Sequence 3, Appli
23	2314	100.0	1290	7	US-10-411-037-67	Sequence 67, Appl
24	2314	100.0	1290	7	US-10-411-026-67	Sequence 67, Appl
25	2314	100.0	1290	7	US-10-410-962-67	Sequence 67, Appl
26	2314	100.0	1290	7	US-10-411-049-67	Sequence 67, Appl
27	2314	100.0	1290	7	US-10-410-930-67	Sequence 67, Appl
28	2314	100.0	1290	7	US-10-410-997-67	Sequence 67, Appl
29	2314	100.0	1290	7	US-10-411-012-67	Sequence 67, Appl
30	2314	100.0	1290	7	US-10-410-913-67	Sequence 67, Appl
31	2314	100.0	1290	8	US-10-851-388-3	Sequence 3, Appli
32	2314	100.0	1290	8	US-10-410-980-67	Sequence 67, Appl
33	2314	100.0	1290	9	US-10-410-897-67	Sequence 67, Appl
34	2314	100.0	1290	9	US-10-984-389-3	Sequence 3, Appli
35	2314	100.0	1296	3	US-09-993-059-9	Sequence 9, Appli
36	2314	100.0	1296	5	US-10-103-327-9	Sequence 9, Appli
37	2314	100.0	1296	7	US-10-602-219-9	Sequence 9, Appli
38	2314	100.0	1296	7	US-10-602-220-9	Sequence 9, Appli
39	2314	100.0	1296	8	US-10-851-388-9	Sequence 9, Appli
40	2314	100.0	1296	9	US-10-984-389-9	Sequence 9, Appli
41	2314	100.0	1306	9	US-10-480-790-5	Sequence 5, Appli
42	2314	100.0	1308	3	US-09-993-059-5	Sequence 5, Appli
43	2314	100.0	1308	5	US-10-103-327-5	Sequence 5, Appli
44	2314	100.0	1308	7	US-10-602-219-5	Sequence 5, Appli
45	2314	100.0	1308	7	US-10-602-220-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-993-059-11
; Sequence 11, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-993-059-11

Alignment Scores:
Pred. No.: 7,21e-278 Length: 1266
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0


```
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAAGGAGAGATTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTCCGCGAGCTA 360
Qy 121 AlaAsnTyrValHisSerIleGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTACAGCAAGAGATGGAAGCTAGGATTTATGCGAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCTGGGAGTTTGGATACACGATGATGATGATGATGATGATGATGAT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGCTTGTACTGTGACAGATTTGGAAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTAAAGCACATGCTTGGCCCTGAAATAGGACTGGGAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGAGTGCATCTTTTAAACAGGAGAGAAATTTGTATGTTGTCTGGACACGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GTTGGAAATGACCCAGATGTTGTATGTTGGCACTTTGGCTCAGCTGGATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGCTAATTCGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGGCAAGCAGGGTACCAGCTTACAGAGGGAGACAACCTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGAGGAGATTGCT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTGCTCTTATACCATCGCAGTGTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrPhe 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAAGCTAGGGTTCATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAGGTTTAAGAAAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
```

RESULT 3

US-10-602-219-11

; Sequence 11, Application US/10602219

; Publication No. US20040016021A1

GENERAL INFORMATION:

; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.

; APPLICANT: Grill, Laurence K.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION

; FILE REFERENCE: LSBC-0087-CP09B

; CURRENT APPLICATION NUMBER: US/10/602,219

; CURRENT FILING DATE: 2003-06-23

; PRIORITY APPLICATION NUMBER: 09/993,059

; PRIORITY FILING DATE: 2001-11-13

; PRIORITY APPLICATION NUMBER: 09/626,127

; PRIORITY FILING DATE: 2000-07-26

; PRIORITY APPLICATION NUMBER: 09/316,572

; PRIORITY FILING DATE: 1999-05-21

; PRIORITY APPLICATION NUMBER: 08/324,003

; PRIORITY FILING DATE: 1994-10-14

; PRIORITY APPLICATION NUMBER: 08/176,414

; PRIORITY FILING DATE: 1993-12-29

; PRIORITY APPLICATION NUMBER: 07/997,733

; PRIORITY FILING DATE: 1992-12-30

; PRIORITY APPLICATION NUMBER: 08/184,237

; PRIORITY FILING DATE: 1994-01-19

; PRIORITY APPLICATION NUMBER: 07/923,692

; PRIORITY FILING DATE: 1992-07-31

; PRIORITY APPLICATION NUMBER: 07/600,244

; PRIORITY FILING DATE: 1990-10-22

; PRIORITY APPLICATION NUMBER: 07/641,617

; PRIORITY FILING DATE: 1991-01-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 11

; LENGTH: 1266

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-10-602-219-11

Alignment Scores:

Pred. No.:	7,21e-278	Length:	1266
Score:	2314.00	Matches:	421
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-10-602-219-12 (1-421) x US-10-602-219-11 (1-1266)

```
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
Qy 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGCTTCTCGGACATCCCTGGGCTAGAGACACTGGCAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCAATGCAACCTTGACTGCAGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGCATCAGTGAGAGCTCTTCATGCAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg 100
Db 241 TGAAGGATGAGGATGAGTATGAGTACCTCTGCTATGATGATGATGATGATGATGATG 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAAGGAGAGATTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTCCGCGAGCTA 360
```

```
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyLeuValAlaAspValGlyAsnLys 140
Db |||
Db 361 GCTAATTATGTTACACGCAAGAGCTGAAGCTAGGAGTTATGACAGATTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db |||
Db 421 ACCTGCGCAGGCTTCCCTGGAGTTTGGATACTACGACATTCATGCTCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db |||
Db 481 GACTGGGAGTAGACTGCTGATAAATTTGATGGTTGTACTGTGACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db |||
Db 541 GCAGATGGTTATAGACACATGCTCTGGCCCTGATAGACTGGCAGAGCAATTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db |||
Db 601 TCTGTGAGTGGCCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgHisPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db |||
Db 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGTCATTTGATGATTCCTGGGAAAGTATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db |||
Db 721 AGTATCTTGAGCTGGACATCTTTTACAGAGAGAGATTTGTGATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db |||
Db 781 GGTGGGAATGACCCAGATATGTTAGTATGGCACTTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db |||
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db |||
Db 901 CGACACATCAGCCCTCAGCCCAAGCTCTCTTCAGGATPAGGACGTAAATGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlnGlyAspAsnPheGluValTrp 340
Db |||
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCAGCTTAGACAGGAGACAACCTTCAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnIleGly 360
Db |||
Db 1021 GAACGACCTCTCTCAGGCTTAGCGTGGCTGTAGCTATATATAAACCGCAGGAGATTGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db |||
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCCTGTATCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db |||
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db |||
Db 1201 TCAAGGTTAAGAGTCACATAAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
QY 421 Met 421
Db |||
Db 1261 ATG 1263
```

RESULT 4

```
US-10-602-220-11
; Sequence 11, Application US/10602220
; Publication NO. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
```

```
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-11
Alignment Scores:
Pred. No.: 7,21e-278 Length: 1266
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0
US-10-602-219-12 (1-421) x US-10-602-220-11 (1-1266)
```

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGACTACATCTGGGCTCGCGCTTGGCTTCGTCTCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyValArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCTCTGGGCTAGAGCCTGGACATGGATTGGCAGAGGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTCAACCTTGACTGCCAGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAGGATGCAAGTTATGAGTACCTCTGCATGTATGACTGTGTGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTACAGCAAGGAGACTGAAGCTAGGAGTTTATGCAATGTTGGAAATAAA 420
```


Db 721 AGTATCTTGGACGTGGACATCTTTTAAACCCAGGAGAGAAATGTTGATGTTGCTGGACACGGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGATGACCCAGATATGATGATGGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCTATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGACGTAATTCCTCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAsnAspPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGATACCACTTTAGACAGGAGACAACCTTTGAAGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTACCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGAGAGCTAGGGTTCATGATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTCACATNAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263

RESULT 6

US-10-984-389-11
; Sequence 11, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; PRIORITY FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-984-389-11

Alignment Scores:

Pred. No.:	7,21e-278	Length:	1266
Score:	2314.00	Matches:	421
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-602-219-12 (1-421) x US-10-984-389-11 (1-1266)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGGCGCTTGGCTTCGCTTCTCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluLeuPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGGATGAGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGCAGACATTCAGGCGAGACCTCAGCGCTTTCCTCATGGGATTGGCCAGTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATATGTTACAGCAAGAGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACAGCAATGATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATCGTTTGTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTATAGCACATGTCCTTGGCCCTGATAGACTGGCAGAGCAATGTTGTATAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGAAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGATGGACATCTTTTAACGAGAGAGAAATGTTGATGTTGCTGGACACGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATTGGCAACTTTGGCACTTGGCCTCAGCTGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCTATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGACGTAATTCCTCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCGCTTAGACAGGAGACAACCTTTGAAGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGAGTTCCTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAGAAGAGCTAGGGTTCTTGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuAsnThr 420
 Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAAACA 1260
 QY 421 Met 421
 Db 1261 ATG 1263

RESULT 7

US-09-993-059-7
 ; Sequence 7, Application US/09993059
 ; Publication No. US20020088024A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GARGER, Stephen A.
 ; APPLICANT: TURPEN, Thomas H.
 ; APPLICANT: KUMAGAI, Monto H.
 ; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
 ; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
 ; FILE REFERENCE: 008010087CPUS06
 ; CURRENT APPLICATION NUMBER: US/09/993,059
 ; CURRENT FILING DATE: 2001-11-13
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 1278
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1275)
 US-09-993-059-7

Alignment Scores:
 Pred. No.: 7,31e-278 Length: 1278
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-09-993-059-7 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTCTGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTCGTTCTCTGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCACTGGGAGGCGCTTCATGTGCAACCTTGGCTGCCAGGAAGGCCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTGTCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGAAGGATGCAGGTATGATGATACCTCTCATTTGATGACTGTGGTGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTGAGAGGAGCAGCTTCAGGCAGACCCCTCAGCGCTTCTCCTCATGGGATTGCCAGGCTA 360

QY 121 AlaIleTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db 361 GCTAATATATGTTTCAGCAAGAGCTAAGCTAGGATTTATGCGAGATGTTTGGAAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGCGCAGCGCTTCCCTGGGAGTTTTGGATACACGACATTTGATGCCAGACCTTTGCT 480
 QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGTAAATTTGATGGTTTACTGTGACAGTTTGGAAATTTG 540
 QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGGTTATAAGCACATGCTTGGCCCTGTAATAGGACTGGCAGAAGCATTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 601 TCCTGTGAGTGGCTCTTTATATATGTCGCTTTTCAAAAGCCCAATATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCATCACTGCGGAAATTTTGTGACATTTGATGATCTCTGGAAGTATATAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACCCAGCAGAGAATTTGTTGATGTTGCTGACCGAG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTGGAAATGACCCAGATATGTTAGTGTATGGCACTTTGGCCCTCAGCTGGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTAGATGGGCTCTGCGCTATCATGCTGCTCTCTTTATTCATGCTCTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAAGCTTTAGACAGGGAGAGCAACTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCCGCGAGGAGTTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGAGTTCCTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAGAAGAGCTAGGGTTCTTGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAAACA 1260
 QY 421 Met 421
 Db 1261 ATG 1263

RESULT 8

US-10-103-327-7
 ; Sequence 7, Application US/10103327
 ; Publication No. US20030106095A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GARGER, Stephen A.
 ; APPLICANT: TURPEN, Thomas H.
 ; APPLICANT: KUMAGAI, Monto H.

;
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-103-327-7

Alignment Scores:
Pred. No.: 7,31e-278 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0
US-10-602-219-12 (1-421) x US-10-103-327-7 (1-1278)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCTCTGGGCC 60
QY 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTCTCTGGGACATCCCTGGGCTAGACACTGGCAATGGATTGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCTGGGAGCGCTTCATGTGCACCTTGACTGCCAGGAAGGCCA 180
QY 61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAGGATGCAGGTTATGAGTACCTCTCATGTGATGACTGTGTGGAGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGACTGAAGTATAGGATTTATGAGATTTATGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTCGGAGTTTGTGATACACGACATTTGATGCTCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTGATGTTGTTACTGTGACAGCTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAACCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCCTCTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240

Db 661 CAGTACTGCAATCACTGGCGAAATTTTGCTGCATTTGATGATTCCTGGAAGATATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACACAGAGAGAATTGTGTATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTAGTTAGTTGGCACTTTGGCCTCAGCTGGAAATCAGCA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGCTATCATGCTGCTCTTTTATTATCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCCAAGCTCTCTTCAGGATAAGGACGTAAATTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGGACAGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCTCTGTGAAGAAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263
RESULT 9
US-10-602-219-7
; Sequence 7, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31

; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-7

Alignment Scores:
Pred. No.: 7,31e-278 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-219-12 (1-421) x US-10-602-219-7 (1-1278)

QY 1 MetGlnLeuArgAspProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTCTGGCC 60
QY 21 LeuValSerTrpAspLeuProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCTCCGGGCTAGAGCACTGGACATGGATGGCAAGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGGCTGCCAGGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAAGAGCTTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGAGATGCAAGTTATGAGTACCTCTGCATGTATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheHisGlyIleArgGlnLeu 120
DB 301 GATTCAGAGGCAGACTTCAGGACAGCCCTCAGCGCTTCTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGATTTATGCAAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATGACTAGCAATGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTGAATCTGCTAAATTTGATGTTGTACTGTGACAGATTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 CGAGATGGTTATAAGCACATGCTTGGCCCTGGATAGGACTGGCAGAGCAATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCAGTGGGCAATTTTGTGACATGATGATTTCTGGAAAGATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

DB 721 AGTATCTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTTGCTGGACCAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
DB 781 GGTGGATGACCCAGATATGTTAGTATTTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCCAAGCTCTCTCTCAGGATAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTTGGGCAAGCAAGGTTACAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCCGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GCACCTCCTCTTATACCATGCAGTTGCTTCCCTGGTAAAGAGTAGGCTGTATCTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAGGAAAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGTTTAAAGAGTACATAAATCCACAGGCACTGTTTGTCTTCAGTAGAAAAACA 1260
QY 421 Met 421
DB 1261 ATG 1263

RESULT 10
US-10-602-220-7
; Sequence 7, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

```
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-7

Alignment Scores:
Pred. No.: 7,31e-278 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-219-12 (1-421) x US-10-602-220-7 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGACCCAGACTACATCTGGGCTGGCGCTTGGCGCTTGGCTTCTTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTTCATGTGCAACCTTGTACTGCCAGGAGGCCA 180

QY 61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTGATCAGTGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTATCAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTCACAGCAAGAGACTGAAGCTAGGAGATTATGACAGATTTGGAATAAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCTCTGGAGTTTGGTACTACGACATTCATGGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTGCTGATAAATTTGATGGTTGTTACTGTACACATTTGGAAAAATTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTCTGAGTGGCTCTTTATATGCGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCAGCTGGCGAATTTTGTCTGACATTCATTCCTCTGGAAAAAGTATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnArgIleValAspValalaglyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGTTCTCGACAGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTTCTCTGGCC 60
```

```
Db 781 GGTGGAAATGACCCAGATATGTTAGTATTGGCAACTTTTGGCCCTCAGCTGGAATCAGCAA 840
QY ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTGGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGGGTACAGCTTACACAGGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGNTAATAAACCCGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTCCTCCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGAAAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTTGCTTCAGTAGAAAAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263

RESULT 11
US-10-851-388-7
; Sequence 7, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUNAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; PRIOR FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-851-388-7

Alignment Scores:
Pred. No.: 7,31e-278 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-10-602-219-12 (1-421) x US-10-851-388-7 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTTCTCTGGCC 60
```

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGAGCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGTCATGTCAGAGAGCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGAGATGCAAGTATGAGTACCTTCGATTTGATGATGCTGGTGGTGGTGGTGGTGGT 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAGGAGAGCTTCAGGAGAGCTTCAGGAGAGCTTCAGGAGAGCTTCAGGAGAGCT 360
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACCTAGCAGATTCAGGAGAGCTTCAGGAGAGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTATGATCTGTAATTTTGTATGTTTGTATGTTTGTATGTTTGTATGTTTGTAT 540
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 541 GCAGATGGTTTATAGACATGTCCTTGGCCCTGAAATAGGAGTGGCAGAGAGCTTCAGGAGAGCT 600
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTTCAAGAGCCCAATATACAGAAATCCGA 660
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGATTCATGATTCCTGGAAGATGATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTACCGAGGAGAGATTTGATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GCTTGGATGACCCAGATATGTTAGTATGGCACTTTGGCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCTCTGGGCTATCATGGCTGCTTCCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 GCACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGCTTAATTCGCAATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGAGGAGGAGTACGAGCTTAGACAGGAGAGCACTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeGly 360
Db 1021 GAACGACCTCTCTGAGCTTAGCTGGGCTAGCTATGATGATGATGATGATGATGATGATGAT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGTAAGGAGTGGCTGTAACTCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTTCATCACAGCTCTCCCTGTGTAAAAAGGAGCTTAGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGTTTAAAGTTCATATAATCCCAAGGAGCTGTTTGTCTCAGTAGAANAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263
RESULT 12
US-10-984-389-7
; Sequence 7, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; PRIORITY FILING DATE: 2004-11-08
; PRIORITY FILING DATE: US/09/993,059
; PRIORITY FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-984-389-7
Alignment Scores:
Pred. No.: 7,31e-278 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-10-602-219-12 (1-421) x US-10-984-389-7 (1-1278)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTGGCTTGGCT 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGTGCATCTGGGAGGCTTCATGTGCAACCTTGCAGGAGAGAGCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGTCATGAGAGAGCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGAGATGCAAGTATGAGTACCTTCGATTTGATGATGCTGGTGGTGGTGGTGGTGGT 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAGGAGAGCTTCAGGAGAGCTTCAGGAGAGCTTCAGGAGAGCTTCAGGAGAGCT 360
QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140

```
Db 361 GCTAATTATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAGATGTTGCAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTCCCTGGGAGTTTGTGATACTACGACATTTGATGCGCCAGACCTTTGCT 480
Qy 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTGCTGAATTTGATGTTGTTACTGTGACAGATTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 CGAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTGAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCATCACTGGCGCAATTTTGTGCAATGATGATTCCTGGAAAGATTAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATTTGTGATGTTGCTGCACAGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTGGAATGACCCAGATATGTAGTATGTCGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGCAAGCAAGGATACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTACCTGGCTGATGCTATGATTAACCGCAGGAGATTTGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCCTGTAATCTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCTTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
```

RESULT 13

```
US-09-993-059-13
; Sequence 13, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
```

```
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (1284)
US-09-993-059-13
```

```
Alignment Scores: 7,36e-278 Length: 1284
Pred. No.: 2314.00 Matches: 421
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 3
```

US-10-602-219-12 (1-421) x US-09-993-059-13 (1-1284)

```
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGGCGCTTGGCTTCGCTTCCTGGCC 60
Qy 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspGlyGluGluPro 60
Db 121 ACCATGGGCTGGCTGCATCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTTATGAGTACCTCTGCATTCATGACTGTGTGATGGCTGCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGAGCAGACTTCAGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCTCTGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
Qy 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTTGTGTTGTTTGTGTCAGCTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGTAGTAGGACTGGCAGAGCAATTTGTGAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCATCACTGGCGCAATTTTGTGCAATGATGATTCCTGGAAAGATTAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
```

Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTGTGGCACTTTGGCCCTGAGCTGGATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaIleProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCTATTGCTAAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTACAGGATAAGGACGTAATTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTGGGCAAGCAGGCTACCGCTTAGACAGGAGACAACTTTGAGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGlnGly 360
Db 1021 GAACGACCTCTCTCAGGCTTACGCTGGGCTGAGCTATGATAAAACCGCAGGATTGGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGAGCTAGGCTTCTATGAGTGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGTTTAAAGAGTCACATAAATCCACAGGCACTGTTTCTTTCAGCTAGAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263

RESULT 14

US-10-103-327-13

; Sequence 13, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; FILE REFERENCE: 008010087CPUS06

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13

; LENGTH: 1284

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; LOCATION: (1)...(1284)

US-10-103-327-13

Alignment Scores:

Pred. No.:	7,36e-278	Length:	1284
Score:	2314.00	Matches:	421
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-10-602-219-12 (1-421) x US-10-103-327-13 (1-1284)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTCGTTCCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGCTTCTCTGGGACATCCCTGGGCTAGACACTGGCAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTCAACCTTGACTGCTGCGAAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGTCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTyrluLysCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGAGGCTTATGAGTACCTCTGCAATGGATTGATGCTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGCAGACTTCAGGCAGACCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACAGCAAGAGGACTGAAGCTAGGATTTATGACAGATGTTGGAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTTGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTGTGTACTGTGACAGCTTTGGAAAATTTG 540
Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpTrpGluIleArg 220
Db 601 TCCTGTAGTGGCCCTCTTTATATGTGGCCCTTTTCAAAAGCCCAATATATACAGAAATCCA 660
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGTGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaIleProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCTATTGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTACAGGATAAGGACGTAATTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTGGGCAAGCAGGCTACCGCTTAGACAGGAGACAACTTTGAGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGlnGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080

Qy	361	GlyProArgSerTyrThrIleAlaValalaSerLeuGlyLysGlyValalaCysAsnPro	380
Db	1081	GGACCTCGCTCTTTATACCATCGCAGTTGCTTCCTGGGTAAAGAGTGCCCTGTAATCCCT	1140
Qy	381	AlaCysPheIleThrGlnLeuLeuProValIysArgLysIeuGlyPheTyrGluTrpThr	400
Db	1141	GCCTCGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGCGGTTCTATGATGGACT	1200
Qy	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr	420
Db	1201	TCAAGTTTAGAGAGTCATAAATCCCACAGGACACTGTTTTGCTTCAGCTAGAAAAACA	1260
Qy	421	Met	421
Db	1261	ATG	1263

RESULT 15

```

US-10-602-219-13
; Sequence 13, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION

```

Alignment Scores:	
Pred. No.:	7,366-278
Score:	2314.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	7
Length:	1284
Matches:	421
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-602-219-12 (1-421) x US-10-602-219-13 (1-1284)

	QY	MetGlnLeuArgAsnProGluHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
	Dd	ATGCAGCTGAAGAACCAGAACTACATCTGGGTGGCGCTTGGCTTCCTTGCC	60

QY	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr	400
Db	1141	GCCTGCTTCATCACAGCTCCTCCCTGTGAAAAGGAGCTAGGGTTCTATGAATGACT	1200
QY	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr	420
Db	1201	TCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAACACA	1260
QY	421	Met	421
Db	1261	ATG	1263

Search completed: December 26, 2005, 15:56:37
Job time : 908 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 10:34:12 ; Search time 4085 Seconds
(without alignments)
4821.876 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MOLRNPELHLCALALRFLA.....RLRSHINPTGVLLQLENTM 421

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10602219/runat_23122005_151142_17149/app_query.fasta_1.583
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10602219 @CEN 1 1 8010 @runat_23122005_151142_17149 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT_DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRSADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hc1:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	1277	4	CR605654
2	2309	99.8	1266	4	CR607242
3	2289	98.9	1253	4	CR617861
4	2104	90.9	1290	10	AY408540 Homo sapi
5	2055	88.8	1290	10	AY408541
6	1811	78.3	1376	4	AK054547 Mus muscu
7	1811	78.3	2962	4	AK040081 Mus muscu

8	1777	76.8	1086	1	AL554978
9	1699	73.4	1014	5	BX354096
10	1691.5	73.1	999	1	AL552630
11	1667.5	72.1	1067	3	BM564282
12	1659.5	71.7	1005	3	BQ062192
13	1649	71.3	922	7	CO645623
14	1648.5	71.2	1026	1	AL577581
15	1632	70.5	1296	10	AY408542 Mus muscu
16	1627.5	70.3	1133	7	CO645672 ILLUMIGEN
17	1589	68.7	920	7	CO645464 ILLUMIGEN
18	1578	68.2	923	7	CT005156
19	1574	68.0	984	7	CO646251 ILLUMIGEN
20	1556.5	67.3	976	5	BQ56043
21	1527	66.0	898	5	BUI91867
22	1519	65.6	922	5	BUI540848
23	1515.5	65.5	937	5	BQ934640
24	1492.5	64.5	852	6	CA454083
25	1465	63.3	888	5	BUI54569
26	1458	63.0	958	6	CA487415
27	1452	62.7	871	6	CA454143
28	1443	62.4	801	5	BUI596617
29	1429	61.8	1025	3	BM450649
30	1425.5	61.6	849	6	CA487531
31	1422.5	61.5	879	2	BG824387
32	1412.5	61.0	927	2	BE622583
33	1411.5	61.0	1071	1	AL575861
34	1408.5	60.9	871	3	BQ225444
35	1406	60.8	908	2	BE379425
36	1404	60.7	796	8	DR156332
37	1402.5	60.6	921	6	CA488935
38	1397	60.4	845	7	CR983967
39	1384	59.8	813	5	BX344841
40	1382	59.7	782	2	B1224248
41	1371.5	59.3	1066	2	BG824323
42	1365	59.0	749	5	BX344842
43	1360	58.8	908	6	CD385342
44	1326	57.3	886	3	BP435628
45	1303	56.3	767	3	B1753664

ALIGNMENTS

RESULT 1

CR605654

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CR605654 1277 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CS0DB008YB11 of Neuroblastoma Cot
10-normalized of Homo sapiens (human).

CR605654 GI:50486461

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605654.1 HTCC

CR605

FEATURES	Location/Qualifiers																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
----------	---------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

```

/plasmid="pCMVSPORT_6"

ORIGIN
Alignment Scores:
Pred. No.:      2,27e-242      Length:      1266
Score:          2309.00        Matches:      420
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    99.78%        Indels:        0
DB:              4            Gaps:          0

US-10-602-219-12 (1-421) x CR607242 (1-1266)

QY      1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB      7 ATGACAGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTTCGCTTCCTGGCC 66

QY      21 LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB      67 CTCGTTTCTCGGGACATCCCTGGGCTAGAGCAGCTGGCAATGGATTGGCAAGGACGCT 126

QY      41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB      127 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 186

QY      61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB      187 GATTCCTGCATCAGTGAGAGGCTTTCATGTGAGATGGCAGAGCTCATGGTCTCAGAGGC 246

QY      81 TrpLysAspAlaGlyTrpGluTrpLysCysIleAspAspCysTrpMetAlaProGlnArg 100
DB      247 TGGAGGATGCAAGGTTATGAGTACCTCTGCAATGTATGATGCTGTGGATGGCTCCCAAGA 306

QY      101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB      307 GATTCAGAGGAGGAGACTTCAGGACAGCCCTCAGCGCTTCTCTCATGGGATTCGCAGGTA 366

QY      121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB      367 GCTAATTTATGTTACAGCAAGGACTGAAGCTAGGATTTATGAGATGTTGGAAATAA 426

QY      141 ThrCysAlaGlyPheProGlySerPheGlyTrpArgPheIleAspAlaGlnThrPheAla 160
DB      427 ACCTCGCAGGCTCCCTCGGGAGTTTGGATACACTACGACATGTATGCCAGACCTTTCCT 486

QY      161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB      487 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 546

QY      181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB      547 GCAGATGGTTATAGCACATGTCCTTGGCCCTGAAATAGGACTGGCAGGAAGCATTTGTGTAC 606

QY      201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
DB      607 TCCTGTGAGTGGCCCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 666

QY      221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAsnSerTrpLysSerIleLys 240
DB      667 CAGTACTGCAATCATCGGCAAAATTTGCTGACATTTGATGATTCCTGGAAAGATATAAG 726

QY      241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB      727 AGTATCTGGACTGCATCTTTTAAACCAGGAGAGAAATTTGTATGTTGCTGGACAGGG 786

QY      261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
DB      787 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 846

QY      281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB      847 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCTATGCTTAATGACCTC 906

```

```

QY      301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB      907 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCCTCAATCAAT 966

QY      321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB      967 CAGGACCCCTTGGCAAGCAGGAGGTACGAGTTAGACAGGAGGAGACACTTTGAGAGTGG 1026

QY      341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB      1027 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATAAAACCGCAGGAGATTGGT 1086

QY      361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB      1087 GGACCTCGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1146

QY      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB      1147 GCCTGCTTCATCACACAGCTCTCTCCTGTGAAAAGGAGAGCTAGGGTTCATGATGGACT 1206

QY      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB      1207 TCAAGGTTAAGAGTCCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAATACA 1266

RESULT 3
CR617861      1253 bp      mRNA      linear      HTC 21-JUL-2004
LOCUS      full-length cDNA clone CS0DK009YE12 of HeLa cells Cot 25-normalized
DEFINITION      of Homo sapiens (human).
ACCESSION      CR617861
VERSION      CR617861.1 GI:50498668
KEYWORDS      HTC; CNSUT_CDNA.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Hominoidea; Homo.
REFERENCE      1 (bases 1 to 1253)
AUTHORS      Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished
REMARK      Contact: Feng Liang Email: fliang@lifetech.com URL:
              http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
              Faraday Avenue
              2 (bases 1 to 1253)
REFERENCE      Genoscope.
AUTHORS      Direct Submission
TITLE      Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
COMMENT      1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
              end enriched, double-strand cDNA was digested with Not I and cloned
              into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
              was normalized. Library was constructed by Life Technologies, a
              division of Invitrogen.
FEATURES      Location/Qualifiers
              source
                1..1253
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS0DK009YE12"
                /tissue type="HeLa cells Cot 25-normalized"
                /plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores:
Pred. No.:      3.53e-240      Length:      1253
Score:          2289.00        Matches:      416
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    98.92%        Indels:        0
DB:              4            Gaps:          0

```

US-10-602-219-12 (1-421) x CR617861 (1-1253)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 6 ATGCAGCTGAGGACCAAGCAATACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 65
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 66 CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGCGCT 125
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 126 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 185
QY 61 AspSerCysIleSerGluLeuValPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 186 GATTCTCGCATCATGTCAGAGAGCTTCCTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 245
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 246 TGGAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 305
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 306 GATTTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 365
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 366 GCTAAATTATGTTTCACAGCAAGAGACTGAAGCTAGGAGATTATGCAGATGTTGGAATAAA 425
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 426 ACCTGGCGAGGCTTCCTGGGAGTTTGGATACACATGATGATGCCAGACCTTGTGT 485
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 486 GACTGGGGAGTAGACTCTGCTAAATTTGATGGTTGTACTGTGACACATTTGGAAAATTG 545
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 546 GCAGATGGTTATGAACACATGCTCTGGCCCTGAATAGGACTGGCAAGACATTTGTATC 605
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 606 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCA 665
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 666 CAGTACTGCAATCACCTGGCGAAATTTTGTCTGACATTTGATGATTCCTGGAAAAGTATAAG 725
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValalaglyProGly 260
Db 726 AGTATCTTGACCTGGACATCTTTTAAACAGGAGAGAAATTGTTGATGTTCTGGACAGGG 785
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 786 GGTGTGAATGACCCAGATATGTAGTATGGCAACTTTGGCTCAGCTGGATCAGCA 845
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 846 GTAACTCAGATGCCCTCTGGGCTATCATGGCTGCTCCTTTATTATCATGTCTAATGACCTC 905
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspIleValIleAlaIleAsn 320
Db 906 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACATAATTCGATCAAT 965
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 966 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAGTGTGG 1025
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1026 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGCT 1085

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1086 GGACCTCCCTCTTATACCATGCAGTTGCTTCCCTGGGTAAGGAGTGGCCGTGTAATCCT 1145
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1146 GCCTGCTTTCATCACACAGCTCTCCCTGTGAAGAAGAGCTAGGCTTCTATGAATGGACT 1205
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGln 416
Db 1206 TCAAGTTAAGAAGTACATAAATCCACAGCACTGTTTGTCTTCAG 1253
RESULT 4
AY408540 1290 bp DNA linear GSS 15-DEC-2003
LOCUS Homo sapiens GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY408540
VERSION AY408540.1 GI:39764511
KEYWORDS GSS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1290
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1290
/gene="GLA"
/locus_tag="HCM3258"

ORIGIN

Alignment Scores:
Pred. No.: 7,95e-220 Length: 1290
Score: 2104.00 Matches: 390
Percent Similarity: 92.64% Conservative: 0
Best Local Similarity: 92.64% Mismatches: 31
Query Match: 90.92% Indels: 0
DB: 10 Gaps: 0

US-10-602-219-12 (1-421) x AY408540 (1-1290)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGACCAAGCAATACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGCGCT 120

Db 181 GATTCTCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCNGAAGGC 240

QY 81 TTPLYAspAlaGlyTyrGluTyrLeuCySIIleAspAspCysTTPMetAlaProGlnAArg 100

Db 241 TGAAGGATGCGAGTTATGAGTACCTCTGCAITGATGACTGTGTGAAGTCTCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnAArgPheHisGlyIleArgGlnLeu 120

Db 301 GATTTCAGAGGCAGACTTCAGGCAGACCTTCAGCGCTTCTCCTCATGGATTCGNAGCTA 360

QY 121 AlaAsnTyrValHisSerIysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140

Db 361 NNNNNNNNGTTTCCACAGCAAGAGCTGAAGCTAGGCAATTTATGCAGATCTTGGAAATAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160

Db 421 ACCTGCGCAGGCTTCCTCGGAGTTTGGATACTACGACATTCATGCCAGACCTTTGCT 480

QY 161 AspTTPGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

Db 481 GACTGGGGAGTAGACTCTGTAATTTGATGGTTGTACTGTGACAGTTTGGAAATTTG 540

QY 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnAArgThrGlyArgSerIleValTyr 200

Db 541 CGCATGNN 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220

Db 601 NNN 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240

Db 661 CAGTACTGCANATCAGCTGGCGAAATTTGTCTGAGTTGATGATCTCTGGAAAGATTAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

Db 721 AGTATCTTGCATCGGACATCTTTTAAACAGGAGAGAAATTTGTGATCTGCTGCAGCAGG 780

QY 261 GlyTrpAsnAspProAspMetIleValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280

Db 781 GGTTGGAATGACCCAGATATGTTAGTGATTTGGCACTTTGGGCTCAGCTGGAAATCAGCA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300

Db 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTCTCTTATCATGCTTAATGACCTC 900

QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320

Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340

Db 961 CAGGACCCCTTGGCGCAAGCAAGGATACCAAGCTTAGACAGNNAGACAACCTTGAAGTGTG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnIleGly 360

Db 1021 GAACGACCTCTCTCAGCTTAGCTGGCTGTAGCTATCATTAACCCGCGAGAGATTGT 1080

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380

Db 1081 GGACCTCGCTCTTATACCATCGCACCTTCTTCCCTCGGTAAAGAGTGGCCTGTAATCCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400

Db 1141 GCCTGTCTTCATCACAGCTCTCCCTGTGAAAGGAAGTAGGGTTCTATGAATGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValIleLeuGlnLeuGluAsnThr 420

Db 1201 TCAAGGTGAAGATCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTAGAAATACA 1260

QY 421 Met 421

Db 1261 ATG 1263

RESULT 6

AK054547

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AK054547 1376 bp mRNA linear HTC 03-APR-2004
Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
full-length enriched library, clone:E330039P08
product:galactosidase, alpha, full insert sequence.

AK054547
AK054547.1 GI:26344324
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 1376)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozawa, T.,
Hori, F., Imotani, K., Iehii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-researc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

Encyclopedia Project of Genome Exploration Research Group in Riken
cDNA library was prepared and sequenced in Mouse Genome

Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>.

FEATURES

source

```

1..1376
/organism="Mus musculus"
/mol_type="mRNA"
/strains="C57BL/6J"
/db_xref="FANTOM.DB:E330039P08"
/db_xref="taxon:10090"
/clone="E330039P08"
/sex="female"
/tissue_type="ovary"
/clone_lib="RIKEN full-length enriched mouse cdna library"
/dev_stage="2 days pregnant adult"
21..1286
/note="unnamed protein product; galactosidase, alpha
(MGI:1347344, GB|NM_013463, evidence: BLASTN, 99%,
match=1373)
putative"
/codon_start=1
/protein_id="BAC35819.1"
/db_xref="GI:26344325"
/translation="MAMKLLSRDTRVCEALPLALVFWSLGVRALDNLGARTPTM
GWLHWKPCNLDCEEPDACSIOELFMQALMVSDGWRDAGDYLCIDCWMPAPR
DSKGRLOAPDFPSGIKHLANYHSGKLGIVADGNKTCAGPFGSGSYDIDAOT
FADWGVLLKEDGCHDSVSLNGYKYMALALNRGSIYVSCWPLYLPPFKPNY
TDIYCYCHMRNFDVDSWESIKNLSWTYYQKEIVVEVAGPWNPDMLVIGNFG
LSWQVQVMTALMAJPAALPLMSLDNRQISSQKALLQNDVIAINOPDKQGYCFR
KENHIEVWERPLSNLAWAVARNLQEIIGPCPYTTQISSLGRGLACNPGCIITQLLPE
KVHLGFYEWTLTKLRVNSPGTVLFRLE"
1358..1363
/note="putative"
1376
/note="putative"

```

polyA_signal

polyA_site

ORIGIN

```

Alignment Scores:
Pred. No.: 1..4e-187 Length: 1376
Score: 1811.00 Matches: 327
Percent Similarity: 88.04% Conservative: 41
Best Local Similarity: 78.23% Mismatches: 50
Query Match: 78.26% Indels: 0
DB: 4 Gaps: 0

```

US-10-602-219-12 (1-421) x AK054547 (1-1376)

```

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 27 ATGAGCTTTTTCAGCAGAGATACGCGGCTGGTCTGTGAGCTTGGCTTGTCCCTGGCT 86
QY 21 LeuValSerTTPAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 87 TTAGTTTTCTCGAGCATTCTTGGGGTCAGAGCATTGGCAATGGCTTGGCGGGACTCCT 146
QY 41 ThrMetGlyTTPLeuHisTTPGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 147 ACTATGGGCTGGCTGCATTGGGAACGTTTCATGTGCACTTGGACTTGGCAAGAAGGCT 206
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 207 GATGCTTCGATTAAGTGAAGCACTTTCATGAGTGGCAGAGCTCATGGTCTCTCATGGC 266
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTTPMetAlaProGlnArg 100
DB 267 TGGCGGATGAGGTTATGACTTCTCATAGTAGACTGTGGTGGCTCCCGAGAGG 326
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 327 GATTCAAGGGGAGGCTTCAGGCAGATCCCAACGCTTCTTCTAGTGGGATCAACACCTC 386

```

```

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 387 GCAAAATACGTCCACAGCAGCAAGAGATTGAAGCTAGGATTATGCGAGATTGTTGGGAATAAA 446
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 447 ACCTGTGCGAGTTTCCCGGGAGTTTTGGATCCTATGACATTTGATGATGATGATGATGATGAT 506
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 507 GACTGGGCGTAGATCTGCTAAATTTTGTGTTGCTACTGTGACAGTGTAGTATCCTTTG 566
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 567 GAGAATGTTTATAAGTACATGCGCTTGGCTTGAACAGGACAGGCGGAGCATTTGTATAC 626
QY 201 SerCysGluTTPProLeuTyrMetTTPProPheGlnLysProAsnTyrThrGluIleArg 220
DB 627 TCCTGTGAGTGGCCACTTTTATTTGAGACCTTTTCAATAGCCCAATATATACAGATATCCNA 686
QY 221 GlnTyrCysAsnHisTTPArgAsnPheAlaAspIleAspAspSerTTPlysSerIleLys 240
DB 687 TATTACTGCATCATTTGGAGAAATTTTCATGATGATGATGATGATGATGATGATGATGATGAT 746
QY 241 SerIleLeuAspTTPThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 747 AATATCTTGTCTTGACAGAGTGTTTTACCAGAGGAGAGATTGTTGAAGTGCCTGCGACGAGC 806
QY 261 GlyTTPAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTTPAsnGlnGln 280
DB 807 AGCTGGAATACCCAGACAGATGTTAGTATGCGCAACTTTGGCTCAGTTGGGACGACGAG 866
QY 281 ValThrGlnMetAlaLeuTTPAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 867 GTGACACAGATGGCCCTCTGGGCTATCATGCGCGCTCCCTACTCATGTCCAACGATCTG 926
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAla 320
DB 927 CGACAAATCAGCTCTCAAGCCAAAGCTCTGCTTCAGAAATAAGGATGTAATTCGCCATCAAC 986
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTTP 340
DB 987 CAAGACCCCTTGGGCAAGCAGCGCTACTGTTTCAGAAAGGAAACCAACATTGAGGTTGG 1046
QY 341 GluArgProLeuSerGlyLeuAlaTTPAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1047 GAACGGCCACTCTCCAACTAGCTGGGCTGTGGCTGTGAGAAACCTCGACGAGATTGGT 1106
QY 361 GlyProArgSerTTPThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1107 GGACCTTGTCTTATACCATCCAGATTTCTTCCCTGGGTAGAGGACTAGCTGCAATCCT 1166
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluThrThr 400
DB 1167 GGCTGCATCATTTACTCAGCTTCTCCCCGAGAAAGTACACCTAGGCTTCTATGAATGGACT 1226
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGlu 418
DB 1227 TTGACCTTAAAACTCGAGTAACCCCTCAGGCACTGTTTGTTCGGTTAGAA 1280

```

RESULT 7

```

AK040081 2962 bp mRNA linear HTC 03-APR-2004
LOCUS Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
DEFINITION library, clone:A430057F16 product:galactosidase, alpha, full insert
sequence.
ACCESSION AK040081
VERSION AK040081.1 GI:26333580
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```

Db      607  TCCCTGAGTGGCCACTTTATTTAGAGACCTTTTATAGCCCAATATACAGATATCCAA 666
Qy      221  GlnTyrCysAsnHisTrpArgAsnPheAlaaspIleAspAspSerTrpLysSerIleLys 240
Db      667  TATTACTGCAATCATTTGGAGAAATTTTGTATGTTTATGATCTCTCTGGAAAGCATAAG 726
Qy      241  SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      727  AATACTTGTCTTGGACAGTGGTTTACCAGAGGAGATTTGTTGAAGTCGTCGACCAAGC 786
Qy      261  GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db      787  AGCTGGAATGACCCAGACATGTTAGTATCGGCAACTTTTGGCCTCAGTTGGGACGACG 846
Qy      281  ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      847  GTGACACAGATGGCCCTCTGGGCTATCATGGCGCTCCCTTACTCATGTCCAAACATCTG 906
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      907  CGCAAAATCAGCTCTCAAGCCAAAGCTCTGCTTTCAGATAAGGATGTAATTCACATCAAC 966
Qy      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      967  CAAGACCCCTTGGCAAGCAGCGCTACTGTTTTCAGAAAGGAAACCAACATTCAGGTTGG 1026
Qy      341  GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeGly 360
Db      1027  GAACGGCCACTCTCAACCTACCCCTGGGCTGTGGCTGTGGAGAAACCTCGCAGAGATTGGT 1086
Qy      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1087  GGACCTTGTCTTATACCATCAGATTTCTCCCTGGGTAGAGACTAGCTTGCATCTCT 1146
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db      1147  GGCTGCATCATTTACTCAGCTTCTCCCGAGAAAGTACACCTAGGCTTCTATGAATGGACT 1206
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGlu 418
Db      1207  TTGACCTTAAAACTCGAGTAACCCCTCAGGCACTGTTTGTTCGGTTAGAA 1260

RESULT 8
AL554978
LOCUS   AL554978 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION
ACCESSION AL554978.3 GI:45859723
VERSION   AL554978
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT  On Feb 15, 2001 this sequence version replaced gi:31276787.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CSODK009BC06QP1&c=10506.r.

```

FEATURES

```

source
1. 1086
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODK009YE12"
/cell_type="HELA"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.: 5,33e-184 Length: 1086
Score: 1777.00 Matches: 342
Percent Similarity: 96.65% Conservatives: 4
Best Local Similarity: 95.53% Mismatches: 10
Query Match: 76.79% Indels: 7
DB: 1 Gaps: 1
US-10-602-219-12 (1-421) x AL554978 (1-1086)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 6 ATGCAGCTGAGGAAACCCAGACTCATCTGGGCTGGCGCTTGC-CTTCGCTTCCTGGCC 64
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 65 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 124
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 125 ACCATGGGCTGGCTGCATCTGGGAGCGCTTCATGTCAACCTTGCCTCCGAGGAGGCCA 184
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 185 GATTCCTGCATCAGTGAGAGCTTTCATGCAGATGGCAGAGCTCATGTCTCAGAAAGC 244
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 245 TCGAAGGATGCAGGTTATGAGTACCTCTGCATTTGATGATGCTGTGGATTGGCTCCCAAGA 304
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 305 GATTCAGAAAGGAGACACTTCAGGACACCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 364
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 365 GCTAATATTATGTTACAGCAAGGACTGAAGCTAGGATTATGCAGATTGTTCGAAATAAA 424
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 425 ACCTCGCAGGCTTCCCTGGGAGTTTGGTACTACTACGACATTCATGATGCCAGACCTTGT 484
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluLeu 180
Db 485 GACTCGGAGTAGATCTGCTAAATTTTATGTTGTTTACTGTGACAGATTGGAAATTTG 544
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 545 GCAGATGTTTATAGCACATGTCCTTGGCCCTGATAGACTGGCAGAGACATTGTGTAC 604
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeArg 220
Db 605 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATATATACAGAAATCCGA 664
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 665 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTCCTCGAAAGATATAAG 724

```

```

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluAtrGileValAspValAlaGlyProGly 260
Db 725 AGTATCTTGGACTGGACATCTTTTAACACGAGGAGAAATTGTGATGTTCTGGACGAGG 784
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 785 GGTGTGAATGACCCAGATATGTTAGTGAITGGCAACTTTGGCCTCAGCTGGAATCAGCAA 844
QY 281 ValThrGlnMetAlaLeuTrpAlaIle-MetAlaAlaProLeuPheMetSerAsnAspLe 300
Db 845 GTAACCTCAGATGGCCCTCTGGGGTATTCATGGCTGCTCTTTATTATCATGCTTAATGACCT 904
QY 300 uArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAs 320
Db 905 CCGACACATCAGCCCTCAGCCAAAGCTCTCTTCAGGATAGGAGCTGATTTGGCATCAA 964
QY 320 nGlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTr 340
Db 965 TCAGGACCCYTK-GGCAAGCAAGGTACAGCTTAGAMAGGGGACAM-TTTGWAAGTGTG 1022
QY 340 pGluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGln 357
Db 1023 GGA-MGA---CTCTCTCAGGTAGCTGGCTGTAGY-ATGATAAACCCGAGG 1069

RESULT 9
BX354096 1014 bp mRNA linear EST 23-APR-2004
LOCUS BX354096 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DC014YH04 5-PRIME, mRNA sequence.
ACCESSION BX354096
VERSION BX354096.2 GI:46550107
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1014)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 5, 2003 this sequence version replaced gi:30379845.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DC014DD02Qp1&c=10506.r.

FEATURES
Location/Qualifiers
1..1014
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC014YH04"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Alignment Scores: 1.8e-175 Length: 1014
Pred. No.: 1699.00 Matches: 317
Score: 1699.00
Percent Similarity: 97.54% Conservative: 0
Best Local Similarity: 97.54% Mismatches: 7

QY 5 AsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAlaLeuValSerTrp 24
Db 31 AACCCAGAAACATACATCTGGGCTGCGGCTTGCSSTT-CGCTTCCTGGCCCTCGTTTCCTGG 89
QY 25 AspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrp 44
Db 90 GACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCCCTACCATGGGCTGG 149
QY 45 LeuHisTrpGluAtrGlyPheMetCysAsnLeuAspCysGlnGluGluProAspSerCysIle 64
Db 150 CTGCACTGGGAGCGCTTCATGTCAACCTTGACTGCCAGGAGCCAGATTCTCTGGCATC 209
QY 65 SerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAspAla 84
Db 210 AGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGCTGGAAGGATGCA 269
QY 85 GlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSerGluGly 104
Db 270 GGTATGAGTACCTCTGCATTGATGACTGTGGATGGCTGCCCAAGAGATTTCAGAAAGGC 329
QY 105 ArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAsnTrpVal 124
Db 330 AGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTAGCTAAATTATGTT 389
QY 125 HisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLysThrCysAlaGly 144
Db 390 CACAGCAAGGACTGAACTAGAGTAGGATTTATGCAGATGTTGGAAATATAAACCTGCGCAGGC 449
QY 145 PheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAlaAspTrpGlyVal 164
Db 450 TTCCCTGGGAGTTTGGATACCTAGCACATTGATGCCAGACTTTCGTGACTGCTGGGAGTGA 509
QY 165 AspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeuAlaAspGly-Ty 184
Db 510 GATCTGCTAAATAATTTGATGGTGTGTACTGTGACAGTTTGGAAATTTGGCAGATGGTATA 569
QY 184 rLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrSerCysGluTr 204
Db 570 TAAGCACATGCTCTTGGCCCTGAATAGACTGGCAGAGCATTTGTGTACTCTCTGTGAGTG 629
QY 204 pProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArgGlnTyrCysAs 224
Db 630 GCCTCTTTATATGTGGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGACAGTACTGCAA 689
QY 224 nHisTrpArgPheAlaAspIleAspAspSerTrpLysSerIleLysSerIleLeuAs 244
Db 690 TCACCTGGCGAAATTTTGTGACATTGATGATTCCTGGAAAGATATAAAGAGTATCTTGA 749
QY 244 pTrpThrSerPheAsnGlnGluAtrGileValAspValAlaGlyProGlyGlyTrpAsnAs 264
Db 750 CTGGACATCTTTTAACACGAGGAGAAATTGTGATGTTGCTGGACGAGGGGTGGAAATGA 809
QY 264 pProAspMetLeuValIleGlyAsnPheGlyLysSerTrpAsnGlnGlnValThrGlnMe 284
Db 810 CCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAATCAGCAAGTAACCTCAGAT 869
QY 284 tAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSe 304
Db 870 GGCCCTCTGGGGCTATCATGGCTGCTCTTTTATTCAATGCTCAATGCTCCGACACATCAG 929
QY 304 rProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspProle 324
Db 930 CCCTCAGGCCAAAGCTCTCTTCAGRTAAGRCGTAAATTGGCCAC-AACWAGGACCCCTK 988
QY 324 uGlyLysGlnGly 328
Db 989 -GGCMAGCAAGGK 1000

```

```

Query Match: 73.42% Indels: 4
DB: 5 Gaps: 0
US-10-602-219-12 (1-421) x BX354096 (1-1014)
QY 5 AsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAlaLeuValSerTrp 24
Db 31 AACCCAGAAACATACATCTGGGCTGCGGCTTGCSSTT-CGCTTCCTGGCCCTCGTTTCCTGG 89
QY 25 AspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrp 44
Db 90 GACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCCCTACCATGGGCTGG 149
QY 45 LeuHisTrpGluAtrGlyPheMetCysAsnLeuAspCysGlnGluGluProAspSerCysIle 64
Db 150 CTGCACTGGGAGCGCTTCATGTCAACCTTGACTGCCAGGAGCCAGATTCTCTGGCATC 209
QY 65 SerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAspAla 84
Db 210 AGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGCTGGAAGGATGCA 269
QY 85 GlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSerGluGly 104
Db 270 GGTATGAGTACCTCTGCATTGATGACTGTGGATGGCTGCCCAAGAGATTTCAGAAAGGC 329
QY 105 ArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAsnTrpVal 124
Db 330 AGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTAGCTAAATTATGTT 389
QY 125 HisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLysThrCysAlaGly 144
Db 390 CACAGCAAGGACTGAACTAGAGTAGGATTTATGCAGATGTTGGAAATATAAACCTGCGCAGGC 449
QY 145 PheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAlaAspTrpGlyVal 164
Db 450 TTCCCTGGGAGTTTGGATACCTAGCACATTGATGCCAGACTTTCGTGACTGCTGGGAGTGA 509
QY 165 AspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeuAlaAspGly-Ty 184
Db 510 GATCTGCTAAATAATTTGATGGTGTGTACTGTGACAGTTTGGAAATTTGGCAGATGGTATA 569
QY 184 rLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrSerCysGluTr 204
Db 570 TAAGCACATGCTCTTGGCCCTGAATAGACTGGCAGAGCATTTGTGTACTCTCTGTGAGTG 629
QY 204 pProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArgGlnTyrCysAs 224
Db 630 GCCTCTTTATATGTGGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGACAGTACTGCAA 689
QY 224 nHisTrpArgPheAlaAspIleAspAspSerTrpLysSerIleLysSerIleLeuAs 244
Db 690 TCACCTGGCGAAATTTTGTGACATTGATGATTCCTGGAAAGATATAAAGAGTATCTTGA 749
QY 244 pTrpThrSerPheAsnGlnGluAtrGileValAspValAlaGlyProGlyGlyTrpAsnAs 264
Db 750 CTGGACATCTTTTAACACGAGGAGAAATTGTGATGTTGCTGGACGAGGGGTGGAAATGA 809
QY 264 pProAspMetLeuValIleGlyAsnPheGlyLysSerTrpAsnGlnGlnValThrGlnMe 284
Db 810 CCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAATCAGCAAGTAACCTCAGAT 869
QY 284 tAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSe 304
Db 870 GGCCCTCTGGGGCTATCATGGCTGCTCTTTTATTCAATGCTCAATGCTCCGACACATCAG 929
QY 304 rProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspProle 324
Db 930 CCCTCAGGCCAAAGCTCTCTTCAGRTAAGRCGTAAATTGGCCAC-AACWAGGACCCCTK 988
QY 324 uGlyLysGlnGly 328
Db 989 -GGCMAGCAAGGK 1000

```

RESULT 10
AL552630
LOCUS
DEFINITION
AL552630 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS01067YJ01 5-PRIME, mRNA sequence.
ACCESSION
AL552630
VERSION
AL552630.3 GI:45857411
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 999)
AUTHORS
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
On Feb 15, 2001 this sequence version replaced gi:31274445.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, Cp 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with NotI and cloned
into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna/s=CS01067YJ01&c=10506.r.
FEATURES
Location/Qualifiers
1..999
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS01067YJ01"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.: 1,17e-174 Length: 999
Score: 1691.50 Matches: 323
Percent Similarity: 98.48% Conservative: 0
Best Local Similarity: 98.48% Mismatches: 5
Query Match: 73.10% Indels: 5
DB: 1 Gaps: 0
US-10-602-219-12 (1-421) x AL552630 (1-999)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 7 ATGCAGCTGAGGACCCAGCACTACATCTGGGCTGGCGCTTGGCGT-CGCTTCTCGGCC 65
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 66 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 125
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 126 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGATGCCAGGAAGGCCA 185
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 186 GATTCTGTCATCTGAGGAGGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 245
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuGlyCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 246 TGGAGGATGCAAGGTATGAGTACCTCTGCAATTGATGACTGTGGATGGCTCCCAAGA 305

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 306 GATTTCAGAAAGGACAGCTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTCCGCCAGCTA 365
QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 366 GCTAATTATGTTTCACGCAAGGAGTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 425
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 426 ACCTGGCAGAGCTTCCCTGGGAGTTTGGATCTACGACATTTGATGCCAGACCTTTGCT 485
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 486 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTTACTGTGACAGTTTGGAAATTTG 545
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 546 GCAGATGGTTATAGACATGCTCTGGCCCTGAATAGGACTGGCAGAGGATTTGTGTAC 605
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
DB 606 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 665
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 666 CAGTACTGCAATCACTGGCGAAATTTTGTGCTGACATTTGATGATTCCTGGAAAGTATAAG 725
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 726 AGTATCTTGGACTGGACATCTTTTAACCAAGAGAGAAATTTGTATGTTGCTGGACCAAGG 785
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 786 GGTGGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGATCAGCA 845
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 846 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 904
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaLeu 320
DB 905 CGACACATCAGCCCTCAGCCCAAGCT-CTCCTTCAGGATAGGACGTAATTTGCAWC-AAT 962
QY 321 GlnAspProLeuGlyLysGlnGly 328
DB 963 CAGGACCCCTTG-GGCAAGCAAGGK 985
RESULT 11
BM564282
LOCUS
BM564282 1067 bp mRNA linear EST 20-FEB-2002
DEFINITION
AGENCOURT_6560104 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742198
5', mRNA sequence.
ACCESSION
BM564282
VERSION
BM564282.1 GI:18811955
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 1067)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

BQ062192 1005 bp mRNA linear EST 02-APR-2000
 AGENCOURT_6829871 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5923640
 5', mRNA sequence.
 BQ062192
 BQ062192.1 GI:19888802
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominoidea; Homo.
 1 (bases 1 to 1005)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Lou Staudt
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2092 row: h column: 09
 High quality sequence stop: 676.
 Location/Qualifiers
 i. .1005
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5923640"
 /tissue_type="lymphoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_99"
 /note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 3,89e-171 Length: 1005
Score: 1659.50 Matches: 312
Percent Similarity: 96.63% Conservatives: 3
Best Local Similarity: 95.71% Mismatches: 5
Query Match: 71.72% Indels: 6
DB: 3 Gaps: 1

US-10-602-219-12 (1-421) x BQ062192 (1-1005)

QY 28 GlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHisTrp 47
DB 1 GGGGCTAGAGCACTGGCAATGGATTGGCAGAGCGCTTACCATTGGGCTGGCTGCACTGG 60
QY 48 GluArgPheMetCysAsnLeuAspCysGlnGluProAspSerCysTleSerGluLys 67
DB 61 GAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCAGATTCTTCATCAGTGAGAG 120
QY 68 LeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAspAlaGlyTrpGlu 87
DB 121 CTCCTTCAGGAGTGGCAGAGCTCATGCTCAGAGGCTGGAAGATGTCAGGTTATGAG 180
QY 88 TyrLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGln 107
DB 181 TACCTCTGATGATGACTGTGTGGATGGCTGCCCAAGAGATTCAAGAGGCAGACTTCAG 240
QY 108 AlaAspProGlnArgPheProHISGlyLeuArgGlnLeuAlaAsnTyrValHisSerLys 127
DB 241 CGAGACCTTCAGCGCTTCCTCATGGGATTCGCCAGTAGCTAAATTATGTTACAGCAAA 300
QY 128 GlyLeuLysLeuGlyLeuTyrAlaAspValGlyAsnLysThrCysAlaGlyPheProGly 147
DB 301 GGACTGAAGTAGGATTTATGCAATGTTGGAATTAACCTGCGCAGGCTTCCTGGG 360
QY 148 SerPheGlyTyrTyrAspIleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLeu 167
DB 361 AGTTTGGATATACACATTGATGCCAGACCTTTGCTGACTGGGGAGTAGATCTGCTA 420
QY 168 LysPheAspGlyCysTyrCysAspSerLeuGluAsnLeuAlaAspGlyTyrLysHisMet 187
DB 421 AAAATTGTGTTGTTACTGTACACAGTTTGGAAATTTGGCAGATGGTTATAAGCACATG 480
QY 188 SerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrSerCysGluTrpProLeuTyr 207
DB 481 TCCTTGGCCCTGANTAGGACTGGCAGAGCATTGTACTCTCTGTGATGGCTCTTTAT 540
QY 208 MetTrpProPheGlnLysProAsnTyrThrGluIleArgGlnTyrCysAsnHisTrpArg 227
DB 541 ATGTGGCCCTTTCAAAAGCCCAATATATACAGAAATCCGACAGTACTGCAATCACTGGCGA 600
QY 228 AsnPheAlaAspIleAspAspSerTrpLysSerIleLysSerIleLeuAspTrpThrSer 247
DB 601 AATTTTGCTGACATTGATTCCTGGAAAGATATAAAGAGATATCTTGACCTGGACATCT 660
QY 248 PheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrpAsnAspProAspMet 267
DB 661 TTTAACAGAGAGAAATTTGATGTTGCTGGACACAGGGGTTGGAATGACCCAGATATG 720
QY 268 LeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGlnValThrGlnMetAlaLeuTrp 287
DB 721 TTAGTGATTGGCAACTTTTGGCTCAGCTGGAATCAGCAAGTAACTCAGATGGCCCTCTGG 780
QY 288 AlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSerProGlnAla 307
DB 781 GCTATCATGGCTGCTCTTTTATCATGTCTAATGACCTCCGACACATCAGCCCTCAAGCC 840

QY 308 LysAlaLeuLeuGlnAspLysAspValIleAla-IleAsnGlnAspProLeu-GlyLysG 327
DB 841 AAAGCTCTCTTCAGGATAAAGAGCTTATTGCCCATCAATCAGGACCCCTTTGGGGCAAGC 900
QY 327 In-GlyTyrGlnLeu-ArgGlnGly-AspAsnPheGlu---ValTrpGluArgProLeuS 345
DB 901 AAAGGTACCAGCTTTAAACAGGGAACCACTTTTGAAGGGTGGGAACGACCCCTCTCT 960
QY 345 erGlyLeu 347
DB 961 CAGGGCTA 968
RESULT 13
LOCUS CO645623
DEFINITION ILLUMIGEN MCQ 24642 Katze MMLV Macaca mulatta cDNA clone
IBIUM-23923 5' similar to Bases 5 to 920 highly similar to human
GLA (Hs.69089), mRNA sequence.
ACCESSION CO645623
VERSION CO645623.1 GI:50567117
KEYWORDS EST.
SOURCE Macaca mulatta (rhesus monkey)
ORGANISM Macaca mulatta
REFERENCE 1 (Bases 1 to 920)
AUTHORS Magness, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agy, M.B.,
Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and
Iadonato, S.P.
TITLE Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human
JOURNAL Genome Biol. 6 (7), R60 (2005)
PUBMED 15398449
COMMENT Contact: C. Magness
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagness@illumigen.com
Sequenced on 2004.03.02. 796 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCCTCACTAAAGGGAACAAA
BACKWARD: CACTATAGGCGAATTGGGTA
Insert Length: 922 Std Error: 0.00
Plate: CL000111 row: E column: 08
Seq primer: CCCTCACTAAAGGGAACAAA
POLYA=No.
FEATURES
Location/Qualifiers
1..922
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBIUM:23923"
/sex="female"
/dev_stage="adult"
/lab_host="Electromax DH10B"
/clone_lib="Katze MMLV"
/note="Organ: liver; Vector: pDONR 222; Site 1: BsrG I;
Site 2: BsrG I; Created from Cloneminer cDNA Library
Construction kit (catalog #18249-029)"
ORIGIN
Alignment Scores:
Pred. No.: 4,86e-170 Length: 922
Score: 1649.00 Matches: 295
Percent Similarity: 98.03% Conservatives: 3
Best Local Similarity: 97.04% Mismatches: 6


```
Query Match: 71.26% Indels: 0
DB: 7 Gaps: 0
US-10-602-219-12 (1-421) x C0645623 (1-922)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 11 ATGCAGCTGAGGAACCCAGAAATGCATCTGGGCTGGCGCTTCTCTCTCTCTCTCTGGCC 70
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 71 CTCGTTCTCTGGGACATCCCTGGGCGCAGAGCACTGCAATGGATGGCCAGGCGCT 130
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 131 ACCATGGCTGGCTGGACCTGGGAGCGCTTCATGTGCACTTGACCTGCCAGGAAGGCCA 190
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 191 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGATGGC 250
QY 81 TrpLysAspAlaGlyTyrgluTyrluLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 251 TCGAAGGATGCAGGTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGCA 310
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnAtgPheProHisGlyIleArgGlnLeu 120
DB 311 GATTTAGAAGGACAGCTTCAGGACAGACCTCAGCGCTTCTCATGGGATTCGCCAGCTA 370
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 371 GCTAATTATGTTACACAGCAAGGACTGAAGCTAGGAGTTATGAGATGTTGGAAATAA 430
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 431 ACCTGCGCAGGCTTCCCTGGAGTTTGGATATCTATGATGATGATGCCAGACCTTTGCT 490
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 491 GACTGGGGAGTAGATCTGCTGAATTTGATGGTTGTTACTGTGACAGTTTGGAAAGTTG 550
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 551 GCAGATGGTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTATC 610
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 611 TCTGTGTGAGTGGCTCTTTATATGTGGCCATTTCAAAGGCCCAATTTACACAGAAATCCGA 670
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 671 CAGTACTGCATACACTGGCGAAATTTTGTGCAATTTGATGATTCCTGGAAAGATATAAG 730
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 731 AGTATCTTGAGCTGGACATCTTTTACCAGGAGAGAAATTTGATGCTTCTGGACCGAGG 790
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheClyLeuSerTrpAsnGlnGln 280
DB 791 GGTGTGAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 850
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 851 GTAACTCAGATGGGCGCTCTGGCTATCATGGCTGCTCTCTTTATTCATGCTAATGACCTC 910
QY 301 ArgHisIleSer 304
DB 911 CGACACATCACC 922
RESULT 14
AL577581/c
LOCUS
DEFINITION AL577581 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
```

```
CDNA clone CS0DK009YE12 3-PRIME, mRNA sequence.
ACCESSION AL577581
VERSION AL577581.3 GI:46256603
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1026)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 16, 2001 this sequence version replaced gi:31315840.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, Cp 5706 - 91057 EVRY cedex - FRANCE
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?e=CS0DK009BC06NP1&c=10506.r.
Location/Qualifiers
1..1026
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK009YE12"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.: 6,5e-170 Length: 1026
Score: 1648,50 Matches: 308
Percent Similarity: 98,41% Conservative: 1
Best Local Similarity: 98,03% Mismatches: 5
Query Match: 71,24% Indels: 2
DB: 1 Gaps: 0
US-10-602-219-12 (1-421) x AL577581 (1-1026)
QY 97 AlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGly 116
DB 955 GCTCCCCCAAGAGATTCAGAACG-AGACTTCAGCA-GACCCCTCAGCGCTTCTCTCATGGG 898
QY 117 IleArgGlnLeuAlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAsp 136
DB 897 ATTCCCGAGCTAGCTTAATTTGTTTACAGCAAGAGACTGAAGCTAGGAGTTATGCGAT 838
QY 137 ValGlyAsnLysThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAla 156
DB 837 GTTGAAATAAAACCTGGCAGGCTTCCCTGGAGTTTGGATACTACGACATTGATGCC 778
QY 157 GlnThrPheAlaAspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSer 176
DB 777 CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAAATTTGATGGTGTGTACTGTGACAGT 718
QY 177 LeuGluAsnLeuAlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArg 196
DB 717 TTGGAAAATTTGGCAGATGTTATAGACATGTCCTTGGCCCTGATAGGAGCTGGCAGA 658
QY 197 SerIleValTyrSerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyr 216
```

```
Db 657 AGCATGTGTACTCTCTGTGAGTGGCCCTCTTATATATGTGGCCCTTCAAAAGCCCAATTAT 598
Qy 217 ThrGluileArgGlnTyrCysAsnHisTrpArgAanPheAlaAaspIleAaspSerTrp 236
Db 597 ACAGAAATCCGACAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGG 538
Qy 237 LysSerIleLysSerIleLeuAaspTrpThrSerPheAanGlnGluArgIleValAaspVal 256
Db 537 AAAAGTATAAAGAGTATCTTGGACTGGACATCTTTTAAACGAGGAGAGAAATTTGTGATGTT 478
Qy 257 AlaGlyProGlyGlyTrpAanAaspProAaspMetLeuValIleGlyAanPheClyLeuSer 276
Db 477 GCTGACACAGGGGGTGGATGACCAAGATGTTAGTGATTTGGCACTTTGGCCCTCAGC 418
Qy 277 TrpAanGlnGlnValThrGlnMetAlaLeuTrpAlaIleMetAlaAapProLeuPheMet 296
Db 417 TGSNATCAGCAAGTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTATG 358
Qy 297 SerAanAaspLeuArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAaspIleAaspVal 316
Db 357 TCTAATGACCTCCGACATCAGCCCTCAAGGCCAAAGCTCTCTCTTCAGGATAAGGACGTA 298
Qy 317 IleAlaIleAanGlnAaspProLeuGlyGlnGlnTyrGlnLeuArgGlnGlyAaspAan 336
Db 297 ATTGCATCAATCAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGAGGAGACAAC 238
Qy 337 PheGluValTrpGluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAanArg 356
Db 237 TTTGAAGTGTGGGAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATTAACCGG 178
Qy 357 GlnGluileGlyGlyProAargSerTyrThrIleAlaValAlaSerLeuGlyGlyVal 376
Db 177 CAGGAGATTGGTGGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGTAAAGGAGTG 118
Qy 377 AlaCysAanProAlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPhe 396
Db 117 GCCTGTAATCTCCCTGCTTCTATCACAGCTCCCTCCMTTGTGAAAGGAAGCTAGGGTTC 58
Qy 397 TyrGluTrpThrSerArgLeuArgSerHisIleAanProThr 410
Db 57 TATGATGGACTTCAAGTTAGAGTAGTAACATAATVCEACA 16
```

RESULT 15

AY408542
LOCUS Mus musculus G1A gene, VIRTUAL TRANSCRIPT, partial sequence, GSS 15-DEC-2003
DEFINITION Genomic survey sequence.

ACCESSION AY408542

VERSION AY408542.1 GI:39764513

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.

TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)

PUBMED 14671302
REFERENCE 2 (bases 1 to 1296)

AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.

TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering

them based on alignment.
FEATURES
Location/Qualifiers
1..1296
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
gene
c1..>1296
/gene="G1A"
/locus_tag="HMC3258"

ORIGIN

Alignment Scores: 6.04e-168 Length: 1296
Pred. No.: 1632.00 Matches: 301
Score: 81.10% Conservative: 38
Percent Similarity: 72.01% Mismatches: 79
Best Local Similarity: 70.53% Indels: 0
Query Match: 10 Gaps: 0
DB:

US-10-602-219-12 (1-421) x AY408542 (1-1296)

Qy 1 MetGlnLeuArgAanProGluLeuHisLeuGlyCysAlaLeuAlaLeuAargPheLeuAla 20
Db 1 ATGAAGCTTTTGACGACAGATACCGGCTGCTGTGAGCTTGGCTTTGCTCCCTGGCT 60
Qy 21 LeuValSerTrpAaspIleProGlyAlaAargAlaLeuAaspAenGlyLeuAlaAargThrPro 40
Db 61 TTAGTTTTCTGGAGCATCTTGGGGTCAGACATTTGGACAATGGCTTGGCGCGGACTCCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAanLeuAaspCysGlnGluLeuPro 60
Db 121 ACTATGGCTGGCTGTCATTGGGAAAGCTTTCATGTGCAACCTTGACTGTCAGAAAGAGCT 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATCGCTGATTAAGTACCACTGTTCATGCAGATGGCAGAGCTCATGGTCTCTGATGGC 240
Qy 81 TrpLysAapAlaGlyTyrGluTyrLysCysIleAaspAenGlyCysTrpMetAlaProGlnArg 100
Db 241 TGGCGGATGCAGGTTATGACTATCTCTGCATAGTACTGTTGGATGGCTCCCGAGAGG 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAaspProGlnArgPheProHisGlyIleAargGlnLeu 120
Db 301 GATTCAAGGCGCAGCTTTCAGGACAGATCCCAAGCTTTCTAGTGGGATCAAAACACCTC 360
Qy 121 AlaAenTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAaspValGlyAanLys 140
Db 361 GCAATTAACGTCACAGCAAGGATTTGAAGCTAGGATTTATGCAGATGTTGGGATATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAaspIleAaspAlaGlnThrPheAla 160
Db 421 ACCTGTGCAGGTTTCCCGGGAGTTTGGATCCTATGACATTTGATGCGCAGACATTTGCT 480
Qy 161 AspTrpGlyValAaspLeuLeuLysPheAaspGlyCysTyrCysAaspSerLeuGluAanLeu 180
Db 481 GACTGGGCGTAGATCTGCTAAAAATTTGATGGTTGTCTACTGTGACAGTAGTAGTATCC 540
Qy 181 AlaAaspGlyTyrLysHisMetSerLeuAlaLeuAanArgThrGlyArgSerIleValTyr 200
Db 541 GAGATGNN 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAenTyrThrGluIleArg 220
Db 601 NNN 660
Qy 221 GlnTyrCysAanHisTrpArgAanPheAlaAaspIleAaspAenSerTrpLysSerIleLys 240
Db 661 TATTACTGCAATCATTTGGAGAAATTTTTCATGATGTTTATGATTCCTGGAAGCATTAAG 720
Qy 241 SerIleLeuAaspTrpThrSerPheAanGlnGluArgIleValAaspValAlaGlyProGly 260
Db 721 AATATCTTGTCTTGGACAGTGTATTACAGGAAGGAGATGTTTGAAGTTCCTGGACCAAGC 780

```
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db |||||
781 AGCTGGAATGACCCAGACATGTTAGTGATCGCAACTTTGGCCTCAGTTGGGACCAGCAG 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db |||||
841 GTGACACAGATGGCCCTCTGGGCTATCATGGCGCTCCCTACTCATGTCCAACGATCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db |||||
901 CGACAATCAGCTCTCAAGCCAAAGCTCTGCTTCAGATAGGATGTAATGCCATCAAC 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db |||||
961 CAAGACCCCTTGGCAAGCAGGGCTACTGTTTCAGAAAGGAAACCCACATTGAGGTTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnIleGly 360
Db |||||
1021 GAACGGCCACTCTCCAACCTAGCCTGGGCTGTGGCTGTGAGAAACCTGCAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db |||||
1081 GGACCTTGTCCTTATACCATCCAGATTTCTCCCTGGGTAGAGGACTAGCCTGCAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db |||||
1141 GGCTGCATCATTTACTCAGCTTCTCCCGAGAAAGTACACCTAGGCTTCTATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGlu 418
Db |||||
1201 TTGACCTTAAAAAATCGAGTAAACCCCTCAGGCACCTGTTTGTTCGGTTAGAA 1254
```

Search completed: December 26, 2005, 13:48:40
Job time : 4103 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 08:34:41 ; Search time 666 Seconds

(without alignments)
4212.966 Million cell updates/sec

Title: US-10-602-219-12

Perfect score: 2314

Sequence: 1 MQLRNPELHGCALALRFLA.....RLRSHINPTGVLLQLENTM 421

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10602219/runat_23122005_151141_17129/app_query.fasta_1.583
-DB=N_Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THRM SCORES=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10602219 @CGN 1.1 1096 @runat_23122005_151141_17129 -NCPU=6 -ICPU=3
-NO MWAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq 21.1*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2314	100.0	1266	6	AAD45223 Human rGA
2	2314	100.0	1266	10	Add84750 Human alp
3	2314	100.0	1266	12	Adj88276 Human WT
4	2314	100.0	1266	12	Adm48680 Human wil

5	2314	100.0	1266	13	ADU66915	Adu66915 Human alp
6	2314	100.0	1266	14	AEA27444	Aea27444 Human alp
7	2314	100.0	1278	6	AAD45221	Aad45221 Human rGA
8	2314	100.0	1278	10	ADD84746	Add84746 Human rGA
9	2314	100.0	1278	12	ADJ88272	Adj88272 Human rGA
10	2314	100.0	1278	12	ADMA8676	Adm48676 Human wil
11	2314	100.0	1278	13	ADU66911	Adu66911 Human alp
12	2314	100.0	1278	14	AEA27440	Aea27440 Human alp
13	2314	100.0	1284	6	AAD45224	Aad45224 Human rGA
14	2314	100.0	1284	10	ADD84752	Add84752 Human alp
15	2314	100.0	1284	12	ADJ88278	Adj88278 Human WT
16	2314	100.0	1284	12	ADMA8682	Adm48682 Human wil
17	2314	100.0	1284	13	ADU66917	Adu66917 Human alp
18	2314	100.0	1284	14	AEA27446	Aea27446 Human alp
19	2314	100.0	1290	6	AAD45219	Aad45219 Human WT
20	2314	100.0	1290	10	ADD84742	Add84742 Human alp
21	2314	100.0	1290	12	ADJ88268	Adj88268 Human WT
22	2314	100.0	1290	12	ADMA8672	Adm48672 Human wil
23	2314	100.0	1290	12	ADMA9739	Adm49739 Human alp
24	2314	100.0	1290	13	ADU74415	Adu74415 Human alp
25	2314	100.0	1290	13	ADU66907	Adu66907 Human alp
26	2314	100.0	1290	14	AEA27436	Aea27436 Human alp
27	2314	100.0	1290	14	AEBA3187	Aeb43187 Human rGA
28	2314	100.0	1296	6	AAD45222	Aad45222 Human rGA
29	2314	100.0	1296	10	ADD84748	Add84748 Human alp
30	2314	100.0	1296	12	ADJ88274	Adj88274 Human WT
31	2314	100.0	1296	12	ADMA8678	Adm48678 Human wil
32	2314	100.0	1296	13	ADU66913	Adu66913 Human alp
33	2314	100.0	1296	14	AEA27442	Aea27442 Human alp
34	2314	100.0	1306	8	ABZ79777	Abz79777 Human alp
35	2314	100.0	1308	6	AAD45220	Aad45220 Human WT
36	2314	100.0	1308	10	ADD84744	Add84744 Human alp
37	2314	100.0	1308	12	ADJ88270	Adj88270 Human WT
38	2314	100.0	1308	12	ADMA8674	Adm48674 Human wil
39	2314	100.0	1308	13	ADU66909	Adu66909 Human alp
40	2314	100.0	1308	14	AEA27438	Aea27438 Human alp
41	2314	100.0	1319	10	ACF80581	Acf80581 Human alp
42	2314	100.0	1339	10	ADH54464	Adh54464 Human alp
43	2314	100.0	1343	2	AAV31801	Aav31801 Nucleotid
44	2314	100.0	1343	3	AAA75434	Aaa75434 Nucleotid
45	2314	100.0	1343	3	AAA70685	Aaa70685 Human alp

ALIGNMENTS

RESULT 1
AAD45223
ID AAD45223 standard; DNA; 1266 BP.
XX
AC AAD45223;
XX
DT 27-DEC-2002 (first entry)
XX
DE Human rGA-8 DNA.
XX
KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGA-8; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1266
FT /tag= a
FT /product= "Human rGA-8 protein"
XX
PN US2002088024-A1.
XX
PD 04-JUL-2002.
XX
PF 13-NOV-2001; 2001US-00993059.
XX
PR 26-JUL-2000; 2000US-00626127.
XX

PA (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 PA (KUMA/) KUMAGAI M H.
 XX
 PI Garger SJ, Turpen TH, Kumagai MH;
 DR WPI; 2002-681656/73.
 DR P-PSDB; AAE28210.
 XX
 PT Novel human alpha-galactosidase polypeptide useful for treating lysosomal
 PT storage diseases.
 XX
 PS Claim 1; Page 42-44; 88pp; English.
 XX
 CC The invention relates to human alpha-galactosidase truncated at the
 CC carboxy terminus and the production of enzymatically active recombinant
 CC human and animal lysosomal enzymes. The invention is useful for producing
 CC lysosomal enzymes for treating lysosomal storage diseases, producing
 CC altered or mutated proteins, enzymatically active or otherwise, to serve
 CC as precursors or substrates for further in vivo or in vitro processing to
 CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-8 DNA
 XX
 SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.15e-237 Length: 1266
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-602-219-12 (1-421) x AAD45223 (1-1266)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCGCTGGCGCTTGGCTTCTCGGCC 60
 QY 21 LeuValSerTrpAspLeuProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTCGTTCTCGGACATCCCTGGGCTAGACCACTGGACATGGATTGGCAAGAGCGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TGAAGAGATGCAAGTTATGATGACCTCTGCATTGATGACTGTGTGATGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTTCAGAGGCGAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 DB 361 GCTAAATTATGTTTCACAGCAAGAGACTGAAGCTAGGGAATTTATGAGATGTTGGAAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTyrAspIleAspAlaGlnThrPheAla 160
 DB 421 ACTTGGCAGGCTTCCCTGGGAGTTTGGATGACTACGACATGATGCCAGACCTTTGCT 480
 QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGACTTTGGAAAATTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 DB 541 GCAGATGGTTATAGCAATGTCCTTGGCCCTGATAGACTGGCAGAGCAATTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 DB 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 DB 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTGATGATTCCTGGAAGATTAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACAGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 DB 781 GGTGGATGACCCAGATATGTTAGTATTGGCAACTTTGGCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCTTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACACATCAGCCCTCAAGCCCAAGCTCTCTCTTCAAGATAGGAGCTAATTTGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 DB 961 CAGGACCCCTTGGGCAAGCAAGGTACCACTTACAGCTTAGACAGGAGACAACCTTTGAACTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValIleMetIleAsnArgGlnGluIleGly 360
 DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGGAGAGATTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 DB 1141 GCCTGCTTCATCACACAGCTCTCTCTGTGAAAAGGAAGCTAGGGTTCCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 DB 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCCTTTCAGCTAGAAAACACA 1260
 QY 421 Met 421
 DB 1261 ATG 1263
 RESULT 2
 ADD84750
 ID ADD84750 standard; DNA; 1266 BP.
 XX
 AC ADD84750;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human alpha-galactosidase rGAL-8 DNA.
 XX
 KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
 KW enzyme replacement therapy; lysosomal disease.
 XX
 OS Homo sapiens.
 XX
 FH Location/Qualifiers
 FT 1..1266
 FT /*tag= a
 FT /product= "Human rGAL-8"
 XX

PN US2003106095-A1.
 XX 05-JUN-2003.
 XX 20-MAR-2002; 2002US-00103327.
 XX 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 XX (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 XX (KUMA/) KUMAGAI M H.
 XX Garger SJ, Turpen TH, Kumagai MH;
 FI WPI; 2003-801257/75.
 DR P-PSDB; ADD84751.
 XX New polynucleotide for producing active recombinant human and animal
 PT lysosomal enzymes in a plant expression system that can be used in enzyme
 PT replacement therapy.
 XX Claim 1; SEQ ID NO 11; 77pp; English.
 XX The invention relates to human alpha-galactosidase derivatives and the
 CC nucleic acids encoding them. The polypeptides are used in a method for
 CC producing active recombinant human and animal lysosomal enzymes in a
 CC plant expression system. The enzymes can be used in enzyme replacement
 CC therapy for the therapeutic treatment of human and animal lysosomal
 CC diseases. This sequence represents DNA encoding a human alpha-
 CC galactosidase derivative polypeptide of the invention.
 XX
 SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.15e-237 Length: 1266
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-602-219-12 (1-421) x ADD84750 (1-1266)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGACCCAGACTACATCTGGCTGGCGCTTGGCTTGGCTTCTGGCC 60
 Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCCTGGACATGGATTGGCAAGGAGCGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCATCTGGGAGGCTTCATGTGCAACCTTGACTGGCAGGAGGCA 180
 Qy 61 AspSerCysIleSerGluLeuLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTCGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
 Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAGGATGAGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTTCAGAGGCAGACTTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 Qy 121 AlaAsnTrpValHisSerIlyGlyLeuLysLeuGlyIleTrpAlaAspValGluAsnLys 140
 Db 361 GCTAATTTATGTTTCACAGCAAGAGCTGAAGCTAGGAGTTTATGTCAGATGTTGGAAATAAA 420
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyTrpAspIleAspAlaGlnThrPheAla 160

Db 421 ACCTGCGCAGGCTTCCCTGGAGTTTTGGATACACGACATTGATGCCACGACCTTTGCT 480
 Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGCTAAANATTTGATGGTTGTACTCTGACAGATTTGGAAATTTG 540
 Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 Db 541 GCAGATGGTTATAAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
 Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAGAGCCCAATATACAGAAATCCGA 660
 Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCATCACTGGCGAAATTTTCTGACATTGATGATTCCTGGAAAGATATAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACCCAGGAGAGATTTGTATGTTGCTGGACAGG 780
 Qy 261 GlyTrpAsnProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTTGGAGATGCCAGATATGTTAGTATTGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTAGATGGCCCTCTGGGCTATCAGTCTGCTCTTTATTCATGCTTAATGACCTC 900
 Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCAGGATAAGGAGCTAATTTGCCATCAAT 960
 Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTAGACAGGGAGACAACCTTTGAAGTGG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTAGTATGATATAACCCGACGAGAGATTGGT 1080
 Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACTCGCTCTTATACCATCGCAGTTCCTTCCCTGGTAAAGGAGTGGCTGTAACTCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAGTAGGGTTCTATGAATGGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCAAGTTTAAAGTACATTAATCCCAAGGAGCTGTTTTGCTTCAGCTAGAAAACACA 1260
 Qy 421 Met 421
 Db 1261 ATG 1263

RESULT 3
 ADJ88276
 ID ADJ88276 standard; DNA; 1266 BP.
 XX
 AC ADJ88276;
 XX 06-MAY-2004 (first entry)
 XX Human WT rGAL-8 (galactosidase) DNA.
 XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
 KW Fabry's disease; Gaucher's disease; human; gene; ds.
 XX Homo sapiens.

XX	Key	Location/Qualifiers	Db		
PH	FT	1. .i266	121	ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCGAGAAAGGCCA	
FT	FT	/*tag= a	QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly
FT	FT	/product= "human WT rGAL-8 protein"	Db	181	GATTTCGTGCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC
XX	XX	US2004016021-A1.	QY	81	TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg
XX	PD	22-JAN-2004.	Db	241	TGGAAGGATGCAGGTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA
XX	XX	23-JUN-2003; 2003US-00602219.	QY	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu
XX	PR	26-FEB-1988; 88US-00160766.	Db	301	GATTCAAGAGCAGACTTCAGGCAGACCTCAGCGCTTCCTCATGGGATTCGCCAGCTA
XX	PR	15-JUL-1988; 88US-00219279.	QY	121	AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys
XX	PR	17-FEB-1989; 89US-00310881.	Db	361	GCTAATATTATGTTACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAA
XX	PR	05-MAY-1989; 89US-00347637.	QY	141	ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla
XX	PR	08-JUN-1989; 89US-00363138.	Db	421	ACCTGGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTGATGCCAGACCTTTGCT
XX	PR	22-OCT-1990; 90US-00600244.	QY	161	AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu
XX	PR	16-JAN-1991; 91US-00641617.	Db	481	GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG
XX	PR	26-JUL-1991; 91US-00737899.	QY	181	AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr
XX	PR	01-AUG-1991; 91US-00739143.	Db	541	GCAGATGGTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC
XX	PR	31-JUL-1992; 92US-00923692.	QY	201	SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg
XX	PR	30-DEC-1992; 92US-00997733.	Db	601	TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA
XX	PR	29-DEC-1993; 93US-00176414.	QY	221	GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys
XX	PR	19-JAN-1994; 94US-00184237.	Db	661	CAGTACTGCAATCACTGGCGAAATTTTGTGCATTTGATGATTCCTCGGAAAGTATAAAG
XX	PR	14-OCT-1994; 94US-00324003.	QY	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly
XX	PR	21-MAY-1999; 99US-00316572.	Db	721	AGTATCTTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGGACACGG
XX	PR	26-JUL-2000; 2000US-00626127.	QY	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln
XX	PR	13-NOV-2001; 2001US-00993059.	Db	781	GSTTGGATGACCCAGATATGTTAGTGTGGCAACTTTGGCCTCAGCTGGAATCAGCAA
XX	PI	(TURP/) TURPEN T H.	QY	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu
XX	PI	(POGU/) POGUE G P.	Db	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTTCATGCTAATGACCTC
XX	PA	(ERWI/) ERWIN R L.	QY	301	ArgHisIleSerProGlnAlaIysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn
XX	PA	(GRIL/) GRILL L K.	Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCCATCAAT
XX	XX	Turpen TH, Pogue GP, Erwin RL, Grill LK;	QY	321	GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp
XX	XX	WPI; 2004-108227/11.	Db	961	CAGGACCCCTTGGGCAAGCAGGGTACCGAGCTTAGACAGGGAGACAACTTTGAAGTGGG
XX	XX	P-PSDB; ADJ88277.	QY	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly
XX	XX	New lysosomal enzymes, useful in treating human and animal lysosomal storage diseases, e.g. Fabry's disease and Gaucher's diseases.	Db	1021	GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT
XX	XX	Claim 1; SEQ ID NO 11; 71pp; English.	QY	361	GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyIysGlyValAlaCysAsnPro
XX	XX	The invention relates to nucleotide encoding galactosidase (GAL). The invention is useful in gene therapy. The polynucleotides and polypeptides are useful in treating human and animal lysosomal storage diseases, e.g. Fabry's disease and Gaucher's diseases. The present sequence is human GAL DNA.	Db	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTTCCCTGGGTAAGGAGTGGCCTGTAATCCT
XX	XX	Alignment Scores:	QY	381	AlaCysPheIleThrGlnLeuLeuProValLysArgIysLeuGlyPheTyrGluTrpThr
XX	XX	Pred. No.:	Db	1141	GCCTGCTTCATCACACAGCTCTCTCCTGTGAAAAGGAAGCTAGGGTTCATGAATGGACT
XX	XX	Score:	1.15e-237	Length: 1266	
XX	XX	Percent Similarity:	2314.00	Matches: 421	
XX	XX	Best Local Similarity:	100.00%	Conservative: 0	
XX	XX	Query Match:	100.00%	Mismatches: 0	
XX	XX	DB:	12	Indels: 0	
XX	XX			Gaps: 0	
XX	XX	US-10-602-219-12 (1-421) x ADJ88276 (1-1266)			
QY		1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20		
Db		1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC	60		
QY		21 LeuValSerTrpAspIleProGlyValArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40		
Db		61 CTCGTTCTCGGACATCTCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT	120		
QY		41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60		

Db 1201 TCAGGTTAAGAGTCACATTAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAAACACA 1260

Qy 421 Met 421
|||

Db 1261 ATG 1263

RESULT 4
ADM48680
ID ADM48680 standard; DNA; 1266 BP.
XX
XX
AC ADM48680;
XX
XX 03-JUN-2004 (first entry)
XX
XX Human wild type rGAL-8 DNA.
XX
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW gene; ds.
XX
XX Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT 1..1266
FT /*tag= a
FT /product= "GAL-8 protein"
XX
XX US2004023281-A1.
XX
XX 05-FEB-2004.
XX
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX (POGU/) POGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
PI P-PSDB; ADM48681.
XX
XX WPI; 2004-142650/14.
XX P-PSDB; ADM48681.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT disease.
XX
XX Disclosure; SEQ ID NO 11; 72pp; English.
XX
XX The present invention relates to novel galactosidase (Gal) proteins such
CC as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
CC of the present invention are useful for producing recombinant lysosomal
CC enzymes for enzyme replacement therapy for treating human and animal
CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human wild type rGAL-8 DNA used in the exemplification of the invention.
XX

SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.15e-237 Length: 1266
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-602-219-12 (1-421) x ADM48680 (1-1266)

Qy 1 MetGlnLeuArgAnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
|||
Db 1 ATGCAGCTGAGGAACCCAGAACTATCATCTGGGCTGCGCGCTTGGCGCTTCGCTTCTTGGCC 60
|||

Qy 21 LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
|||
Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCATGGACATGGATGGATGGCAAGGACGCT 120
|||

Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
|||
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCACTGCAACCTTGACTGCGCAGGAAGGCA 180
|||

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
|||
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
|||

Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
|||
Db 241 TGGAAGGATGAGGTATGAGTACCTCTGCAATTGATGACTGTGGATGGCTCCCAAGA 300
|||

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyVileArgGlnLeu 120
|||
Db 301 GATTCAGAGGCGACATTCAGCGAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
|||

Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
|||
Db 361 GCTAATTATGTTTCAGCAAAAGGACTGAAGCTAGGAGTTTATGCGAGATGTTGGAAATAA 420
|||

Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
|||
Db 421 ACCTGGCGAGGCTCCCTGGGAGTTTTGGATACTACGACATGATGCGCCAGACCTTGGCT 480
|||

Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
|||
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTTACTCTGCAGCTTGGAAAAATTTG 540
|||

Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|||
Db 541 GCAGATGGTTATAAGCACATGCTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
|||

Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
|||
Db 601 TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCCAATTATACAGAAATCCGA 660
|||

Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|||
Db 661 CAGTACTGCATCACTGCGGAAATTTTGTGCATTTGATGATTCCTGGAAGATATAAG 720
|||

Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
|||
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTATGTTGTGTGGACAGGG 780
|||

Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|||
Db 781 GGTTGGAGTACCAGATATGTTAGTGTAGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
|||

Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
|||
Db 841 GTAACCTAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
|||

Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|||

Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCCTCAAT 960
 Qy 321 GlnAspProLeuGlyGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTACGCTTGGGCTGAGCTATGATAAACCGCAGGAGATTGT 1080
 Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGCCCTGTAATCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACAGCTCCTCCTGTGTAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
 Qy 421 Met 421
 Db 1261 ATG 1263
 RESULT 5
 ID ADU66915 standard; DNA; 1266 BP.
 XX AC ADU66915;
 XX DT 10-FEB-2005 (first entry)
 XX DE Human alpha-galactosidase protein encoding DNA #5.
 XX KW Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
 KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
 KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
 KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human; gene; ds.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 XX FT 1..1266
 XX FT /*tag= a
 XX FT /product= "Human alpha-galactosidase protein"
 XX US2004234516-A1.
 XX PN 25-NOV-2004.
 XX PD 21-MAY-2004; 2004US-00851388.
 XX PF 26-JUL-2000; 2000US-00626127.
 XX PR 13-NOV-2001; 2001US-00930305.
 XX PR 20-MAR-2002; 2002US-00103327.
 XX PA (LARG-) LARGE SCALE BIOLOGY CORP.
 XX PI Garger SJ, Turpen TH, Kumagai MH;
 XX DR WPI; 2004-821274/81.
 XX DR P-PSDB; ADU66916.
 XX PT A pharmaceutical composition comprising a lysosomal enzyme, useful for
 PT enzyme replacement therapy for the treatment of lysosomal storage
 PT diseases, such as Fabry's disease.
 XX PS Disclosure; SEQ ID NO 11; 88pp; English.
 XX CC The present invention relates to the production of human and animal

CC lysosomal enzymes in plants by a transient plant expression system. The
 CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
 CC galactosidase enzymes having a post-translational modification provided
 CC by the plant expression system. The invention is useful in enzyme
 CC replacement therapy for treating lysosomal storage diseases such as
 CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
 CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
 CC also useful in researches for developing new approaches to medical
 CC treatment of lysosomal storage diseases and in industrial processes
 CC involving enzymatic substrate hydrolysis. The present sequence is the
 CC human alpha-galactosidase protein encoding DNA.
 XX SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
 Alignment Scores: 1.15e-237 Length: 1266
 Pred. No.: 2314.00 Matches: 421
 Score: 2314.00 Conserv: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 100.00% Gaps: 0
 DB: 13
 US-10-602-219-12 (1-421) x ADU66915 (1-1266)
 Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGGCGCTTGGCTTCCTCTGGCC 60
 Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGAGAGCCA 180
 Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTGTCATCAGTGAAGACTCTTCATGAGATGGCAGAGCTCATGGTCTCAGNAGGC 240
 Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGGATTGGCTCCCAAGA 300
 Qy 101 AspSerGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTGAGAGGACAGACTTCAGGACACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGAGATTATGCAGATTGTTGGAATATA 420
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCAGGCTTCCTGGCAGTTTTTGGATACTACGACATTGATGCCAGACCTTTGCT 480
 Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTTGTACTGTGACAGATTTGGAAATTTG 540
 Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGGTTATAGCACATGTCTTGGCCCTGATAGGACTGGCAGAGCAATTGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 601 TCCTGTAGTGGCTCTTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
 Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTCAATCACTGGCGAAATTTTCTGACATTGATGATTCCTCGAAAGATATAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

Db 721 AGTATCTTGGACTGGACATCTTTAAACAGAGAGAAATGTGTGTTGCTGGACCAAGG 780
 Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTGGATGACCCAGATATGTAGTATGGCACTTTGGCCTCAGCTGGAAACGCA 840
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTCAGATGGCCCTCTGGCTATCATGCTGCTCTTTATTTCATGCTTAATGACCTC 900
 Qy 301 ArgHisIleSerProGlnAlaValAlaLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGGCATCAAT 960
 Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGNCCCTTGGCAGCAAGGTACCACTTAGACAGGAGACNACTTTGAAGTGTGG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCAGCAGGAGATGGT 1080
 Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGCATGTTGCTTCCCTGGGTAAAGGAGTGCCCTGTAATCCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCTCTCTCTGTAAGAGAGCTAGGGTTCATGATGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCAAGGTAAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAAACA 1260
 Qy 421 Met 421
 Db 1261 ATG 1263
 RESULT 6
 ID AEA27444 standard; DNA; 1266 BP.
 AC AEA27444;
 XX
 DT 11-AUG-2005 (first entry)
 XX
 DE Human alpha-galactosidase DNA, rGAL-8, SEQ ID NO: 11.
 XX
 KW Gauchers disease; metabolic; neurological disease; niemann pick disease;
 KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
 KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 CDS 1..1266
 FT /*tsa= a
 FT /product= "Human alpha-galactosidase protein"
 XX
 US2005125859-A1.
 FN
 XX
 PD 09-JUN-2005.
 XX
 PP 08-NOV-2004; 2004US-00984389.
 XX
 PR 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 PR 20-MAR-2002; 2002US-00103327.
 XX
 PA (LARG-) LARGE SCALE BIOLOGY CORP.

PI Garger SJ, Turpen TH, Kumagai MH;
 DR WPI; 2005-404004/41.
 DR P-PSDB; AEA27445.
 XX
 PT New isolated polypeptides useful for producing lysosomal enzymes in
 PT plants to be utilized in enzyme replacement therapy or for the
 PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
 PT Gaucher's disease.
 XX
 PS Disclosure; SEQ ID NO 11; 88pp; English.
 CC
 CC The present invention relates to the production of human and animal
 CC lysosomal enzymes in plants by a transient plant expression system. The
 CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
 CC galactosidase (Gal) enzymes having a post-translational modification
 CC provided by the plant expression system. The invention is useful in
 CC enzyme replacement therapy for treating lysosomal storage diseases such
 CC as Gaucher's disease, Niemann-pick disease, Fabry's disease, Tay-Sachs
 CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
 CC also useful in researches for developing new approaches to medical
 CC treatment of lysosomal storage diseases and in industrial processes
 CC involving enzymatic substrate hydrolysis. The present sequence is the
 CC human alpha-galactosidase DNA.
 XX
 SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,15e-237 Length: 1266
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservations: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 14 Gaps: 0
 US-10-602-219-12 (1-421) x AEA27444 (1-1266)
 Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
 Qy 21 LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCATGGACAATGGATTGGCAAGGACGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGCTGGCTGGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCAGGAAGAGCA 180
 Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTGCATCAGTGAGNAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 Qy 81 TrpLysAspAlaGlyTrpGluTyrlieuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAAAGGATGAGGTTATGAGTACCTCTGCACTGATGATGATGATGATGATGATGAT 300
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTTCAGAGGACACTTCAGGAGACCTCAGGCTCTTCATGGGCTTCTCATGGATTCGCCAGCTA 360
 Qy 121 AlaSerTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
 Db 361 GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGATTTATGAGATGTTTGGAAATAAA 420
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACACGATTCATGATGATGATGATGATGAT 480
 Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGCTAATAATTTGATGGTTTACTGTGACAGTTTGGAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
Db |||
541 GCAGATGGTTATTAACACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 600
QY 201 SerCysGluTyrProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db |||
601 TCCTGTGAGTGGCCTCTTATATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaSerIleAspAspSerTrpLysSerIleLys 240
Db |||
661 CAGTACTGCAATCACTGGCGAAATTTTGCTGCATTTGATGATTCCTGGAAAGATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspAlaValProGly 260
Db |||
721 AGTATCTTGGACTGGACATCTTTTAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db |||
781 GGTGGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaIaProLeuPheMetSerAsnAspLeu 300
Db |||
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db |||
901 CGACACATCAGCCCTCAAGCCAAAGCTTCCTTCAGGATAGGAGTAAATTCCTCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db |||
961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnIleGly 360
Db |||
1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATTAACCGGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db |||
1081 GGACCTCGCTCTTATACCATCGAGTGTCTCCCTGGGTAAAGGAGTGGCCCTGTAAATCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db |||
1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db |||
1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTTCCTTCAGCTAGAAAAACACA 1260
QY 421 Met 421
Db |||
1261 ATG 1263
RESULT 7
ID AAD45221 standard; DNA; 1278 BP.
XX AAD45221;
AC AAD45221;
XX 27-DEC-2002 (first entry)
DT Human rGAL-4 DNA.
DE Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-4; gene; ds.
XX Homo sapiens.
OS Key Location/Qualifiers
FH 1..1278
FT /*tag= a
FT /product= "Human rGAL-4 protein"
XX US2002088024-A1.
PN

XX 04-JUL-2002.
XX 13-NOV-2001; 2001US-00993059.
XX 26-JUL-2000; 2000US-00626127.
XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUNA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2002-681656/73.
DR P-PSDB; AAE28208.
XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
PT storage diseases.
PS Claim 1; Page 36-38; 88pp; English.
XX The invention relates to human alpha-galactosidase truncated at the
CC carboxy terminus and the production of enzymatically active recombinant
CC human and animal lysosomal enzymes. The invention is useful for producing
CC lysosomal enzymes for treating lysosomal storage diseases, producing
CC altered or mutated proteins, enzymatically active or otherwise, to serve
CC as precursors or substrates for further in vivo or in vitro processing to
CC a specialised industrial form for research or therapeutic uses, to
CC produce more effective therapeutic enzyme, for producing antibodies
CC against lysosomal enzymes for medical diagnostic use, and in any
CC commercial process that involves substrate hydrolysis. The present
CC sequence is human rGAL-4 DNA
XX SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,17e-237 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-602-219-12 (1-421) x AAD45221 (1-1278)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db |||
1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db |||
61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACCTGGACAAATGGATTGGCAAGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db |||
121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCACACCTTGACTGCCAGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db |||
181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db |||
241 TGGAAAGGATGACGGTTATGAGTACCTCTGCATTTATGATCTGTGTATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db |||
301 GATTGAGAAGGCAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db |||
361 GCTAATTATTTTACAGCAAGAGACTGAAGCTAGGAGTTTATGCGAGATTTTGCAGATGTTGGAAATAAA 420

```
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGAGGCTTCCCTGGAGTGTGGATACAGCAATGTATGATGCCAGACCTTGTCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGATTGGAAAAATTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 CGAGATGGTTATAAGCACATGTCCTTGGCCCTGATAGACTGGCAGGAAGCAATGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTAGTGGCCCTCTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCGGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATGTATGATTCCTGGAAAAGTATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTGGACTGACATCTTTTAACCCAGGAGAGAAATGTTGATGTTGCTGGACACGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTAGTATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGTAAGGACGTAAATGGCAATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGGAGACAACTTTGAAAGTGTG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTGCTTCCCTGGGTAAAGGAGTGGCCCTGTAATCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGTTTAAAGAAGTCACATAAATCCACAGGCACTGTTTGTCTTTCAGCTAGAAAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263
RESULT 8
ADD84746
ID ADD84746 standard; DNA; 1278 BP.
XX
AC ADD84746;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human alpha-galactosidase rGAL-4 DNA.
XX
KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease.
XX
```

```
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 1..1278
FT /*tag= a
FT /product= "Human rGAL-4"
XX
PN US2003106095-A1.
XX
PD 05-JUN-2003.
XX
PF 20-MAR-2002; 2002US-00103327.
XX
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
Garger SJ, Turpen TH, Kumagai MH;
XX
WPI; 2003-801257/75.
P-FSDB; ADD84747.
XX
New polynucleotide for producing active recombinant human and animal
PT lysosomal enzymes in a plant expression system that can be used in enzyme
PT replacement therapy.
XX
PS Claim 1; SEQ ID NO 7; 77pp; English.
XX
CC The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents DNA encoding a human alpha-
CC galactosidase derivative polypeptide of the invention.
SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.17e-237 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-602-219-12 (1-421) x ADD84746 (1-1278)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGCCTTCCTCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyValAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGTGGCAGGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGly 80
Db 181 GATTCCTGCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAGGATGCAAGTTATGATGACTCTGCAATTGATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
XX
```

```

Db      301 GATTGAGAGGAGAGCTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy      121 AlaAsnTyrValHisSerIysGlyLeuIysLeuGlyIleTyrAlaAspValGlyAsnIys 140
      |||
Db      361 GCTAATTATGTTACAGCAAGAGCTGAAGAGCTAGGAGATTATGCGAGATTGCGAATAAAA 420
Qy      141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
      |||
Db      421 ACCTGGCAGGGCTTCCTGGGAGTTTGGATACTAGCAATGTATGCCAGACCTTTGCT 480
Qy      161 AspTrpGlyValAspLeuLeuIysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
      |||
Db      481 GACTGGGAGTAGACTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTTG 540
Qy      181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
      |||
Db      541 GCAGATGGTTATTAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAAGCAATGTGTAC 600
Qy      201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
      |||
Db      601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy      221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpIysSerIleIys 240
      |||
Db      661 CAGTACTGCAATCACTGGCGAAAATTTGCTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Qy      241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
      |||
Db      721 AGTATCTTGGACTGGACATCTTTTAACACGAGAGAAATTTGTTGATGTTGCTGGACAGGG 780
Qy      261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
      |||
Db      781 GGTGGGATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGGAATCAGCAA 840
Qy      281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
      |||
Db      841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTCATACGACCTC 900
Qy      301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
      |||
Db      901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
Qy      321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
      |||
Db      961 CAGGACCCCTTGGGCAAGCAAGGTACCACTTAGACAGGAGACAACTTTCAGTGTGG 1020
Qy      341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
      |||
Db      1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCAGGAGAGATTGGT 1080
Qy      361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
      |||
Db      1081 GGACCTCGCTCTTATACCATCGCAGTGTCTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
Qy      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
      |||
Db      1141 GCTGTCTTATCACAGCTCTCCCTCTGTGAAAGGAAGCTAGGTTCTATGATGGACT 1200
Qy      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
      |||
Db      1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260
Qy      421 Met 421
      |||
Db      1261 ATG 1263

RESULT 9
ID ADJ88272
AC ADJ88272 standard; DNA; 1278 BP.
XX
XX ADJ88272;
XX
XX 06-MAY-2004 (first entry)

```

```

XX      Human rGAL- 4 (galactosidase) DNA.
DE
XX      Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW      Fabry's disease; Gaucher's disease; human; gene; db.
XX
OS      Homo sapiens.
XX
XX      Location/Qualifiers
FH      CDS
FT      1..1278
FT      /*tag= a
FT      /product= "Human WT rGAL-4 protein"
FT      /partial
FT      /note= "No start codon"
XX
XX      US2004016021-Al.
XX
XX      22-JAN-2004.
XX
XX      23-JUN-2003; 2003US-00602219.
XX
XX      26-FEB-1988; 88US-00160766.
XX      26-FEB-1988; 88US-00160771.
XX      15-JUL-1988; 88US-00219279.
XX      17-FEB-1989; 89US-00310881.
XX      05-MAY-1989; 89US-00347637.
XX      08-JUN-1989; 89US-00363138.
XX      22-OCT-1990; 90US-00600244.
XX      16-JAN-1991; 91US-00641617.
XX      26-JUL-1991; 91US-00737899.
XX      01-AUG-1991; 91US-00739143.
XX      31-JUL-1992; 92US-00923692.
XX      30-DEC-1992; 92US-00937733.
XX      29-DEC-1993; 93US-00176414.
XX      19-JAN-1994; 94US-00184237.
XX      14-OCT-1994; 94US-00324003.
XX      21-MAY-1999; 99US-00316572.
XX      26-JUL-2000; 2000US-00626127.
XX      13-NOV-2001; 2001US-00930599.
XX
XX      (TURP/) TURPEN T H.
PA      (POGU/) POGUE G P.
PA      (ERWI/) ERWIN R L.
PA      (GRIL/) GRILL L K.
XX
PI      Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX      WPI; 2004-108227/11.
DR      P-PSDB; ADJ88273.
XX
XX      New lysosomal enzymes, useful in treating human and animal lysosomal
PT      storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX      Claim 1; SEQ ID NO 7; 71pp; English.
XX
XX      The invention relates to nucleotide encoding galactosidase (GAL). The
CC      invention is useful in gene therapy. The polynucleotides and polypeptides
CC      are useful in treating human and animal lysosomal storage diseases, e.g.
CC      Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC      DNA.
XX
XX      Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;
SQ

```

```

Alignment Scores:
Pred. No.:      1.17e-237      Length:      1278
Score:          2314.00      Matches:      421
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              12      Gaps:      0

```

US-10-602-219-12 (1-421) x ADJ88272 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTGCATCAGTCAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysValleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAAAGGATGCAAGGTTATGAGTACTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAAGAGCAGACTTCAGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaSerTrpValHisSerIleGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 361 GCTAATTATGTTACAGCAAGAGCTGAAGCTAGGAGTATTATGCAGATGTTGGAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTTGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTG 540
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 541 GCAGATGTTTATAAGCACATGCTCTGGCCCTGAAATAGACTGGCAGAACATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
DB 601 TCCTGTAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCATCATCTGGCGAAATTTTGTGCATTTGATGATTCCTGGAAAGATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGAAATGACCCAGATATGTAGTGTAGTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGAGCTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGCAAGCAAGGTTACCAGCTTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTACCCCTGGGCTGTAGCTATGATAAACCCGCGAGAGATTGGT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380

DB 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAGAAGAGCTAGGGTTCTATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGGTTAAGAGTTCACATAAATCCACAGGCACCTGTTTGTCTTCACTAGAAAAACA 1260
QY 421 Met 421
DB 1261 ATG 1263
RESULT 10
ADM48676
ID ADM48676 standard; DNA; 1278 BP.
XX
AC ADM48676;
XX
XX 03-JUN-2004 (first entry)
XX
DE Human wild type rGAL-4 DNA.
XX
KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 4..1278
FT /*tag= a
FT /product= "GAL-4 protein"
FT /partial
FT /note= "No start codon"
XX
PN US2004023281-A1.
PD
XX 05-FEB-2004.
PF
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX (POGU/) POGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
PI Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-142650/14.
DR P-PSDB; ADM48677.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT disease.

XX Disclosure; SEQ ID NO 7; 72pp; English.

XX The present invention relates to novel galactosidase (Gal) proteins such

CC as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions

CC of the present invention are useful for producing recombinant lysosomal

CC enzymes for enzyme replacement therapy for treating human and animal

CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick

CC disease, Fabry disease and Tay-Sachs disease. The present sequence is

CC human wild type rGAL-4 DNA used in the exemplification of the invention.

XX

SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,17e-237 Length: 1278

Score: 2314.00 Matches: 421

Percent Similarity: 100.00% Conservatives: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 12 Gaps: 0

US-10-602-219-12 (1-421) x ADM48676 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20

DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40

DB 61 CTCGTTCCTGGGACATCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60

DB 121 ACATGGGCTGGCTGCATCTGGAGGCTTCATGTGCACCTTGACTGCCAGGAAGGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80

DB 181 GATTCTCTGATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100

DB 241 TGAAGGATGCAAGGTATAGTACCTCTGCATGTATGACTGTGGATGGCTCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

DB 301 GATTCAAGGACAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140

DB 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGCAGATGTTGGAAATAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160

DB 421 ACCTGCGCAGGCTTCCTCGGAGTTTGTGATATCTAGCATTTGATGTCGCCAGCCTTGT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

DB 481 GACTGGGAGTAGAATCTGTAATAATTGATTTGTTACTGTGACACTTTGGAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200

DB 541 CGAGATGGTTATAAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrTrpGluIleArg 220

DB 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240

DB 661 CAGTACTGCANACACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCGGCAGGAGTTGT 1080

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380

DB 1081 GGACCTCGCTCTTATACCATGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400

DB 1141 GCCTGCTCATCATCAGAGCTCCTCTGTGAAAGAGAGTAGGTTCTATGAATGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420

DB 1201 TCAAGTTAAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCCTCAGTAGAAACACA 1260

QY 421 Met 421

DB 1261 ATG 1263

RESULT 11

ADU66911

ID ADU66911 standard; DNA; 1278 BP.

XX AC ADU66911;

DT 10-FEB-2005 (first entry)

XX DE Human alpha-galactosidase protein encoding DNA #3.

XX Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;

KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;

KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;

KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human; gene; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 1..1278

FT /tag= a

FT /product= "Human alpha-galactosidase protein"

FT /transl_except= (pos:1..6, aa:Gln)

XX PN US2004234516-A1.

XX PD 25-NOV-2004.

XX PF 21-MAY-2004; 2004US-00851388.

XX PR 26-JUL-2000; 2000US-00626127.

PR 13-NOV-2001; 2001US-00993059.

PR 20-MAR-2002; 2002US-00103327.

XX (LARG-) LARGE SCALE BIOLOGY CORP.

XX

PI Garger SJ, Turpen TH, Kumagai MH;

DR WPI; 2004-821274/81.

DR P-PSDB; ADU66912.

XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.

XX Disclosure; SEQ ID NO 7; 88pp; English.

XX The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCS, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein encoding DNA.

XX SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,17e-237 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-10-602-219-12 (1-421) x ADU66911 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTCTCTGGGACATCCCTGGGCTAGAGCACTGGAATGGGATGGCAAGGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCATCGGAGGCGCTTCATGTGCAACCTTGACTGCAGGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTGAGAGGACAGACTTCAGGACAGCCCTCAGCGCTTTCCTCATGGGATGGCCAGCTA 360
QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 361 GCTAATTATGTTACAGCAAGAGGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACAGCAATTCATGATGCCAGACCTTGTCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTGACAGATTGGAAATTG 540
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200

DB 541 GCAGATGGTTATAAGCAGCATGTCTCTGGCCCTGTAATAGGACTGGCAGAAAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
DB 601 TCCGTGTAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTTGATGATTCCTGCGAAAGTATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATTTGTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGATGATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAAATCAGNA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTTATTTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACATCATCAGCCCTCAAGCCCAAGCTCTCTTCAGGATAGGACGTAAATGGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTGGGCAAGCAAGGTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATATAAACCGCAGGAGATTGT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCTCTCTGTGAAAGGAGCTAGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGGTTAAGAGTACATATAATCCCAAGGAGCTGTTTGTCTTTCAGCTAGAAAAACA 1260
QY 421 Met 421
DB 1261 ATG 1263
RESULT 12
ID AEA27440 standard; DNA; 1278 BP.
XX AEA27440;
AC AEA27440;
DT 11-AUG-2005 (first entry)
XX Human alpha-galactosidase DNA, rGAL-4, SEQ ID NO: 7.
DE Gauchers disease; metabolic; neurological disease; niemann pick disease;
KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
KW antileptic; cns-gen.; lysosome storage disease; alpha-galactosidase;
KW gene; ds.
XX Homo sapiens.
XX Location/Qualifiers
FH 1..1278
FT CDS
FT /*tag= a
FT /product= "Human alpha-galactosidase protein"
FT CDS 4..1278

FT /*tag= b
FT /product= "human alpha-galactosidase protein"
FT /partial
FT /note= "No start codon"
XX

US2005125859-A1.

09-JUN-2005.

08-NOV-2004; 2004US-00984389.

26-JUL-2000; 2000US-00626127.

13-NOV-2001; 2001US-00993059.

20-MAR-2002; 2002US-00103327.

(LARG-) LARGE SCALE BIOLOGY CORP.

Garger SJ, Turpen TH, Kumagai MH;

WPI; 2005-404004/41.

P-PSDB; AEA27441, AEA27472.

XX New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.

PS Disclosure; SEQ ID NO 7; 89pp; English.

CC The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase (Gal) enzymes having a post-translational modification
CC provided by the plant expression system. The invention is useful in
CC enzyme replacement therapy for treating lysosomal storage diseases such
CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase DNA.

SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.17e-237 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-602-219-12 (1-421) x AEA27440 (1-1278)

QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
Db	1	ATGCAGCTGAGGACCCAGACTACATCTGGGCTGGCGCTTGGCGTTCCTCTGGGC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
Db	61	CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGCGCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	121	ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCACCTTGACTGCCAGGAGGCCA	180
QY	61	AspSerCysIleSerGluLeuLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	181	GATTCTGTCATCATGTCAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGNAGGC	240
QY	81	TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg	100
Db	241	TGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA	300

AC AAD45224;
 XX 27-DEC-2002 (first entry)
 XX DE Human rGAL-8R DNA.
 XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
 KW therapeutic; rGAL-8R; gene; ds.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT CDS 1..1284
 FT /*tag= a
 FT /product= "Human rGAL-8R protein"
 XX PN US2002088024-A1.
 XX PD 04-JUL-2002.
 XX PP 13-NOV-2001; 2001US-00993059.
 XX PR 26-JUL-2000; 2000US-00626127.
 XX (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 PA (KUMA/) KUMAGAI M H.
 XX Garger SJ, Turpen TH, Kumagai MH;
 XX WPI; 2002-681656/73.
 DR P-PSDB; AAE28211.
 XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
 PT storage diseases.
 XX Claim 1; Page 44-46; 88pp; English.
 XX The invention relates to human alpha-galactosidase truncated at the
 CC carboxy terminus and the production of enzymatically active recombinant
 CC human and animal lysosomal enzymes. The invention is useful for producing
 CC lysosomal enzymes for treating lysosomal storage diseases, producing
 CC altered or mutated proteins, enzymatically active or otherwise, to serve
 CC as precursors or substrates for further in vivo or in vitro processing to
 CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-8R DNA
 XX Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 1.17e-237 Length: 1284
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-602-219-12 (1-421) x AAD45224 (1-1284)

Oy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACTCTGGGCTGGCGCTTGCCTTCTCGGCC 60
 Oy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTGCTTTCTTGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120
 Oy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCACTGGAGGCGCTTCATGTGTCAACCTTGACTGCCAGGAAGAGCCA 180

Oy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTCGCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 Oy 81 TrpLysAspAlaGlyTrpGluTyrlleuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAAGGATGAGGTTATGAGTACCTCTGCATTGATGACTTGTGGATGGCTCCCAAGA 300
 Oy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTCAAGGAGGAGACTTCAGCGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 Oy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
 Db 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGGAATTATGCAGATGTTTGGAAATAA 420
 Oy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCAGCGCTTCCCTGGGAGTTTGGATACACGACATTTGATGCCAGACCTTTTGT 480
 Oy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
 Oy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 Db 541 GCAGATGGTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAAGCAATTGTGTAC 600
 Oy 201 SerCysGluTrpProLeuTyrlleuMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
 Db 601 TCCTGTGAGTGGCCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCA 660
 Oy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCATCACTGGCGAAATTTTCTGCACATTGATGATTCCTGGAAAGATATAAG 720
 Oy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTTGTCTGGACAGGG 780
 Oy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTGGGAATGACCCAGATATGTTAGTATGGCACTTTTGGCCCTCAGCTGGGAATCAGCAA 840
 Oy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTTAATGACCTC 900
 Oy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTGGCCATCAAT 960
 Oy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTAGACAGGGAGAGCAACTTTGAAGTGTGG 1020
 Oy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCCGAGGAGATTGGT 1080
 Oy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
 Oy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAGAGCTAGGGTCTTATGAATGGACT 1200
 Oy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCAAGGTTAAGAAGTACATAAATCCCAACAGGCACTGTTTTGTCTTCAGCTAGAAAACACA 1260

```

QY      421 Met 421
DB      1261 ATG 1263

RESULT 14
ID      ADD84752
XX      ADD84752 standard; DNA; 1284 BP.
XX      ADD84752;
XX      29-JAN-2004 (first entry)
XX      Human alpha-galactosidase rGAL-8R DNA.
XX      Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
XX      enzyme replacement therapy; lysosomal disease.
XX      Homo sapiens.
XX      Key Location/Qualifiers
XX      CDS 1..1284
XX      /tag= a
XX      /product= "Human rGAL-8R"
XX      US2003106095-A1.
XX      05-JUN-2003.
XX      20-MAR-2002; 2002US-00103327.
XX      26-JUL-2000; 2000US-00626127.
XX      13-NOV-2001; 2001US-00993059.
XX      (GARG/) GARGER S J.
XX      (TURP/) TURPEN T H.
XX      (KUNAI/) KUMAGAI M H.
XX      Garger SJ, Turpen TH, Kumagai MH;
XX      WPI; 2003-801257/75.
XX      P-PSDB; ADD84753.
XX      New polynucleotide for producing active recombinant human and animal
XX      PT lysosomal enzymes in a plant expression system that can be used in enzyme
XX      PT replacement therapy.
XX      Claim 1; SEQ ID NO 13; 77pp; English.
XX      The invention relates to human alpha-galactosidase derivatives and the
XX      CC nucleic acids encoding them. The polypeptides are used in a method for
XX      CC producing active recombinant human and animal lysosomal enzymes in a
XX      CC plant expression system. The enzymes can be used in enzyme replacement
XX      CC therapy for the therapeutic treatment of human and animal lysosomal
XX      CC diseases. This sequence represents DNA encoding a human alpha-
XX      CC galactosidase derivative polypeptide of the invention.
XX      SQ Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.17e-237 Length: 1284
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-602-219-12 (1-421) x ADD84752 (1-1284)

QY      1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB      1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCCTGCTGCGCC 60
QY      21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40

```

Db 1141 GCCTGCTTCATCAGACAGCTCCCTCTGTGAAAGAGAGCTAGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGCTTCAGCTAGAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
RESULT 15
ADJ88278
ID ADJ88278 standard; DNA; 1284 BP.
XX
AC ADJ88278;
XX 06-MAY-2004 (first entry)
XX Human WT rGAL-8 R (galactosidase) DNA.
XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease; human; gene; ds.
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 1..1284
FT CDS /*tag= a
FT /product= "Human WT rGAL-8R protein"
XX
XX US2004016021-Al.
XX
XX 22-JAN-2004.
XX
XX 23-JUN-2003; 2003US-00602219.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 15-JUL-1988; 88US-00219279.
XX 17-FEB-1989; 89US-00310881.
XX 05-MAY-1989; 89US-00347637.
XX 08-JUN-1989; 89US-00363138.
XX 22-OCT-1990; 90US-00600244.
XX 16-JAN-1991; 91US-00641617.
XX 26-JUL-1991; 91US-00737899.
XX 01-AUG-1991; 91US-00739143.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.

(TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.

Turpen TH, Pogue GP, Erwin RL, Grill LK;

WPI; 2004-108227/11.
P-PSDB; ADJ88279.

New lysosomal enzymes, useful in treating human and animal lysosomal storage diseases, e.g. Fabry's disease and Gaucher's diseases.

Claim 1; SEQ ID NO 13; 71pp; English.

The invention relates to nucleotide encoding galactosidase (GAL). The invention is useful in gene therapy. The polynucleotides and polypeptides are useful in treating human and animal lysosomal storage diseases, e.g.

CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL DNA.
XX
SQ Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1..17e-237 Length: 1284
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-602-219-12 (1-421) x ADJ88278 (1-1284)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyValAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGAGCGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGCATCAGTGAGNAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGNAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAGGATGCAGGTTATGAGTACCTCTGCTGATGACTGTTGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGCAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTAATGTTTCAGCAAGAGACTGAAGTAGGATTTATGCAATGTTGGAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGTGATACACGACATTTGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTTGATGGTTGTACTGTGCACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGTCTTGGCCCTGAATAGACTGGCAGAGCAATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTCGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTGGACTGGACATCTTTTAACCCAGAGAGAAATTTGTATGTTGTGTGACACAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTGATTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300

```
Db      841  GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTATCATGTCTAATGACCTC 900
Qy
301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      901  CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
Qy      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      961  CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGGAGACAACTTTGAGTGTGG 1020
Qy      341  GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021  GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGCAGGAGATTGGT 1080
Qy      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081  GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db      1141  GCCTGCTTCATCACACAGCTCCTCCTGTGAAGAGGAGCTAGGGTTCTATGAATGGACT 1200
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db      1201  TCAAGGTTAAGAAAGTCACATAAATCCCA CAGGCACCTGTTTTTGCCTTCAGCTAGAAAACACA 1260
Qy      421  Met 421
Db      1261  ATG 1263
```

Search completed: December 26, 2005, 11:08:34
Job time : 683 secs